Demand Estimation in Airline Revenue Management

A dissertation presented to the

School of Business and Economics
of
Freie Universität Berlin

in candidacy for the degree of
doctor rerum politicarum
(Dr. rer. pol.)

by
Dipl.-Wi.-Ing. Philipp Bartke

2014

Referees

- Univ.-Prof. Dr. Natalia Kliewer
- Univ.-Prof. Dr. Catherine Cleophas

Publications

- Bartke, P., Cleophas, C., & Zimmermann, B. (2013). Complexity in airline revenue management. *Journal of Revenue and Pricing Management, 12*(1), 36–45.
- Bartke, P., & Cleophas, C. (2013). *Data Driven Parameterization of Simulation Models for Airline Revenue Management.* Department of Information Systems, Freie Universität Berlin, Germany. Working Paper.
- Bartke, P., Cleophas, C., & Kliewer, N. (2013). *Airline Revenue Management as a Control Problem – New Demand Estimates.* Department of Information Systems, Freie Universität Berlin, Germany. Working Paper. Submitted to *European Journal of Operations Research.*

Date of Defense: February 11, 2014

Declaration of Authorship

Except where reference is made in the text of this thesis, this thesis contains no material published elsewhere or extracted in whole or in part from a thesis presented by me for another degree or diploma. No other person's work has been used without due acknowledgement in the main text of the thesis. This thesis has not been submitted for the award of any other degree or diploma in any other tertiary institution.

Date *Signature*

Zusammenfassung

Ziel des Airline Revenue Managements ist es, den Umsatz einer Fluggesellschaft aus kurzfristiger, taktischer Sicht zu maximieren. Durch die gezielte Verfügbarkeitssteuerung einzelner Preispunkte sollen Auslastung und mittlerer Erlös pro Fluggast optimiert werden. Die Vielzahl möglicher Reisewege durch das Netzwerk einer großen Fluggesellschaft, die große Zahl unterschiedlicher Buchungsklassen und die Tatsache, dass Kunden bis zu einem Jahr im Voraus reservieren können, führen dazu, dass Fluggesellschaften die Verfügbarkeit von Hunderten von Millionen Preispunkten steuern müssen. Dies ist daher nur mit Hilfe von mathematischen Optimierungsmodellen und weitreichender IT-Unterstützung möglich.

Ein Großteil der Revenue Management Optimierungsmodelle aus der Literatur setzt voraus, dass ein Modell der Marktnachfrage existiert und bekannt ist. In der Praxis jedoch ist dies nicht der Fall: die Nachfrage muss aus beobachteten Verkaufs- und Verfügbarkeitsdaten geschätzt werden. Die Qualität dieser Nachfrageschätzung ist dabei häufig ausschlaggebend für die Gesamtleistung eines Revenue Management Systems.

In dieser Arbeit werden die Probleme der Nachfrageschätzung und der Umsatzoptimierung als ein Zustandsraum-Modell formuliert, und es wird gezeigt, wie existierende Schätzmethoden für solche Modelle für das Airline Revenue Management nutzbar gemacht werden können. Darüber hinaus wird die Berechnung der Cramér-Rao Schranke für dieses Modell erläutert, die eine untere Schranke für den mittleren quadrierten Schätzfehler, ungeachtet der verwendeten Schätzmethode, darstellt. Eine Simulationsstudie zeigt, dass die vorgeschlagenen Methoden größtenteils besser abschneiden als bestehende Schätzverfahren und einen Schätzfehler aufweisen, der nahe an der Cramér-Rao Unterschranke liegt.

Mit komplexer werdenden Revenue Management Systemen steigt die Zahl der Parameter des Nachfragemodells, während die Gesamtzahl der Buchungen der Größenordnung nach konstant bleibt. Wir zeigen, dass für jene Parameter, die das Kundenwahlverhalten beschreiben, jede Schätzmethode einem beliebig großen mittleren quadrierten Schätzfehler unterliegt, wenn die Zahl der Buchungen, auf denen die Schätzung basiert, gegen null tendiert. Mit Hilfe einer Simulation wird dieses theoretische Ergebnis bekräftigt und ein Umsatzverlust von etwa 1,5% durch diesen "Effekt der kleinen Zahlen" gemessen.

Um dem Kleine-Zahlen-Effekt entgegenzutreten, wird in dieser Arbeit eine Methode zur Verschmelzung von Nachfrageprognosen vorgeschlagen, die auf den strukturellen Eigenschaften der Cramér-Rao Schranke basiert und das Wissen über die Unsicherheit in der aktuellen Nachfrageprognose ausnutzt. Eine Simulationsstudie zeigt, dass die vorgeschlagene Prozedur den oben genannten Umsatzverlust in großen Teilen vermeiden kann.

Abstract

The objective of airline revenue management is the maximization of an airline's revenue on a tactical, short-term level. This is achieved by controlling the availability of individual posted prices to find an optimal trade-off between utilization and average yield per passenger. The large number of potential itineraries through a large airline's network, the number of different booking classes and the fact that tickets are sold up to one year in advance imply that airlines need to control the availability of hundreds of millions of such price points. This creates the necessity for mathematical optimization models and pervasive IT support.

A majority of revenue management optimization models in the literature assumes that some model of market demand is known. In practice however, this model is in fact never known and has to be estimated from observed bookings and availabilities, the quality of these estimates being crucial for the overall performance of the revenue management system. We formulate the demand estimation and revenue optimization problem as a state-space model and illustrate how existing and well-known state-space estimation methods can be adapted for the airline revenue management problem. Moreover, we describe how to compute the Cramér-Rao bound for the estimation problem which provides a lower bound on the mean-squared estimation error of any estimation procedure. In a simulation study we show that one of the proposed methods compares favorably to existing approaches in most cases and features an estimation error that is close to the theoretical lower bound.

As revenue management systems become increasingly sophisticated, the number of parameters in the demand model grows while the total number of booking events remains roughly constant. Specifically for parameters describing customer choice, we show that any estimation procedure must exhibit an arbitrarily large mean-squared estimation error as the number of booking events that the estimate is based on tends to zero. Simulation results confirm this theoretical result and predict an overall revenue loss of about 1.5% due to this effect.

Finally, we propose a so-called forecast merging procedure which makes use of the structural properties of the Cramér-Rao bound and exploits information about the uncertainty of current demand estimates provided by our proposed estimation method. A simulation study shows that the merging procedure can mitigate the negative revenue effect described above to great extent.

Contents

List of Figures

List of Tables

List of Symbols and Acronyms

Note: Included are symbols and acronyms that are used across more than one subsection.

Symbols

A	History of past availability vectors
B	History of past booking vectors
$C_a(x_{choice})$	Choice Model, giving the probability vector of an arriving customer to buy each of the offered products
$H_a(x)$	Demand model, giving expected bookings for availability vector a and demand parameter vector x
I_t	Fisher information matrix at time t, see equation (4.15)
$MSE^A_{choice,1}$	Mean-squared error of choice parameters for merging cluster A *before* merging, see equation (8.4)
$MSE^A_{choice,2}$	Mean-squared error of choice parameters for merging cluster A *after* merging, see equation (8.5)
M_t	Fisher measurement information matrix, describing the information gain from observing sales at time t, see equation (4.17)
$N(\mu, \Sigma)$	Multi-variate Normal or Gaussian distribution with mean μ and covariance matrix Σ
P_t	Covariance matrix for the uncertainty in demand estimate x_t
P_{choice}	Covariance matrix for the customer choice parameters, a diagonal block of P

P_{cross}	Cross-covariance matrix between the demand volume and customer choice parameters, an off-diagonal block of P
P_{vol}	Covariance matrix for the demand volume parameters, a diagonal block of P
$Poi(\lambda)$	Poisson distribution with arrival rate λ
Q	Covariance matrix of the state-evolution error term w_t, see equation (4.9)
R	Covariance matrix for the observation error term v_t, see equation (4.13)
Δ^A_{MSE}	Difference between mean-squared errors of choice parameters for merging cluster A before and after merging, see equation (8.11)
Δ^x_y	$(\nabla_x)(\nabla_y)^T$, if $x = y$ this is the Hessian matrix, page 34
∇_x	$\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n}$, i.e. the gradient operator, page 34
a	Availability vector
b	Vector of current bookings or sales
$diag(d)$	Diagonal matrix with diagonal vector d
f	Vector of prices or fares for each product
i	Index, usually into the vector of sold products
t	Index of the time period, usually $1 \leq t \leq T$
$tr(M)$	The trace of matrix M, i.e. the sum of its diagonal elements
v_t	Error term in the observation or measurement equation, see equation (4.13)
w_t	Error term in the state-evolution equation, see equation (4.9)
x	Demand parameter vector
x_{choice}	Part of demand parameter vector that describes customer choice
x_{vol}	Part of demand parameter vector that describes overall demand volume

Acronyms

AE	Active Estimation, see section 6.3
EM	Expectation-Maximization
FP	Forecast Prediction, see section 6.4.1
MLE	Maximum-Likelihood Estimation, see section 6.4.1
MMSE	Minimum mean-squared error
O&D	Origin-destination pair
PCRB	Posterior Cramér-Rao (lower) bound, see section 4.4
PF	Particle Filter, see section 6.2
PODS	Passenger Origin-Destination Simulator
SE	Simple Estimation, see section 6.4.1
UKF	Unscented Kalman Filter, see section 6.1

Part I.

Motivation

1. Introduction to Revenue Management

The goal of revenue management is the maximization of a company's profits on a tactical, short- to medium-term level. Variable costs are commonly assumed to be negligible, such that profit maximization is equivalent to revenue maximization; hence, the term revenue management. The potential of revenue management is greatest when there is customer heterogeneity in the market, such that prices can be set individually for different demand segments. Figure 1.1 depicts the value of customer segmentation graphically: achieved revenue is given by the gray area, which becomes larger when an additional price point is used to sell to a subset of customers.

In the airline industry, for example, there are two main demand segments: business and leisure travelers. Customers, especially those from the high-value business segment, have little incentive to reveal which segment they belong to. Therefore, airlines use a variety of mechanisms to prevent business customers purchasing products intended for the leisure segment. These so-called fences or restrictions include advance purchase rules, itinerary restrictions and stricter cancellation policies for leisure products. While these restrictions decrease product value for all customer segments, the assumption is that leisure customers are more willing to purchase their ticket early, stay longer at the destination and need not rebook nor cancel their ticket; business travelers, in contrast, will prefer the more expensive business products in the face of these conditions. This mechanism will never segment the market perfectly and moreover it alienates some segments of demand completely, e.g. leisure travelers who want to book late, stay only for a day, or need more flexibility. However, it has served the airline industry well in the past, as we will see in the example of American Airlines.

Sales structure to facilitate market segmentation and posted price points are usually determined in the medium-term, adjusting to changing market conditions and competitor behavior. The short-term, tactical component arises when production capacity is fixed in the medium-term while demand fluctuates. Then, the availability of cheaper

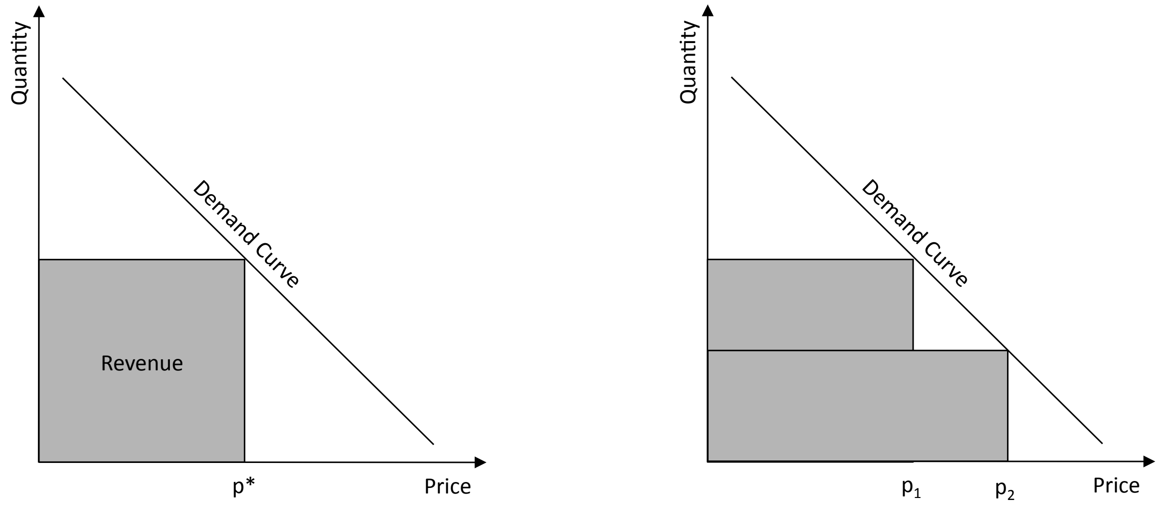

Figure 1.1.: Graphical representation of additional revenue generated by selling at different prices to separate customer segments (Adapted from Talluri & Ryzin 2005)

products has to be carefully controlled, in order to avoid stock-outs before demand for the higher priced products has been exhausted. Forecasting the demand to come for different products and deducing the optimal availability of all products is the purpose of revenue management in a stricter sense and is also the focus area of this thesis.

The stochastic nature of demand and the interdependence of different demand segments via the capacity limit create the need for automated decision support systems in the practice of revenue management. The first such system in wide-spread use was the Dynamic Inventory Allocation and Maintenance Optimizer (DINAMO), which went live in 1985 at American Airlines (Talluri & Ryzin 2005, p. 9). In the preceding years, American Airlines had already introduced new fare products directed at leisure travelers, the so-called (Ultimate) Super-Saver Fares with strong restrictions but low prices, which often undercut all competitors' prices, even those of the newly established low-cost carrier PeopleExpress. These new fares were now quantity controlled by DINAMO, such that more Super-Saver tickets could be sold on flights with low utilization and only few or none would be sold on flights that had high utilization from business travelers alone. This strategy proved to be highly successful, generating about $700 million in additional revenue from dynamic quantity control for American Airlines between 1988 and 1991 (Smith et al. 1992), and forced PeopleExpress into bankruptcy in less than two years.

Today, revenue management is considered an essential practice for every airline, providing about 4-5% additional revenue, which is in many cases about as large as the whole profitability margin (Talluri & Ryzin 2005, p. 10). Many other service industries start adopting revenue management practices as well, including hotels, cruise lines, car rental companies, theme parks, etc.

1.1. Complements and Substitutes

For the most part, early revenue management systems assumed a one-to-one relationship between products and resources, i.e. each sold product corresponds to one unit of used capacity on a single resource. In many industries, however, this is not the case. Airlines for example, sell multiple connecting flights on a single ticket, hotels sell multiple consecutive nights in a room. From the customers' point of view, the multiple services that make up such a compound product have complementary values: the customer is only willing to purchase the complete bundle, or nothing at all.

This quality links the revenue management decision between multiple resources in intricate ways. The availability of a low fare product on a short-haul flight may strongly influence sales on a connecting long-haul flight. Network-based revenue management methods have emerged to explicitly address these interdependencies from complementary products. Talluri & Ryzin (2005, p. 82) report expected revenue gains of 1.5-3% from network-based revenue management. However, this comes at the cost of increased complexity, in both forecasting and optimization. Demand has now to be forecasted for all sold resource combinations, increasing the number of forecasted entities by orders of magnitude. Now, an airline has to forecast demand for every possible itinerary through its network, instead of just forecasting the aggregated demand per flight. Massively more data has to be processed, and often data sparsity issues arise, since many origin-destination combinations are only rarely sold.

The complexity of the optimization problem also increases drastically. While exact and computationally efficient algorithms exist to find the optimal seat allocation for a single resource, this is no longer true for the multi-resource case. In practice, airlines usually resort to heuristic decomposition strategies, where the network problem is decomposed into a number of single resource problems. Nevertheless, the potential revenue gains had already prompted 38% of the airlines surveyed by Weatherford (2009) to convert to network-based revenue management systems.

As outlined by Dunleavy & Phillips (2009), low-cost carriers operate with far fewer fare restrictions than traditional carriers, often relying solely on the time of purchase for price differentiation. While greatly diminishing their ability to segment the market, this allows them to sell to customer segments that were previously excluded from the market by too many fare restrictions. In the face of this competition, many traditional carriers reconsider the heavy use of fare restrictions on their products. Especially on short- to medium-haul markets, where competition with low-cost carriers is strong, traditional carriers are increasingly removing their fare restrictions to keep their products competitive.

Traditional revenue management systems assumed that demand for each product was independent of the demand for other products. However, the lack of strong demand segmentation lets customers choose more freely between the different products that an airline offers for a single origin-destination combination. These products are perfect substitutes from the customers' point of view, i.e. a customer will purchase at most one of the products in her consideration set. This choice behavior of customers poses serious issues to traditional revenue management systems: Cooper et al. (2006) show that forecast quality will degrade and revenues will decrease over time. This is known as the spiral-down effect.

To avoid spiral-down and re-align the revenue management model with reality, choice-based revenue management systems have been developed that model the choice behavior of customers explicitly. Fiig et al. (2009) show that existing optimization procedures can still be used for choice-based revenue management when properly transforming the inputs. The main challenge is therefore in forecasting. Now, not only the volume of customers per product has to be estimated, but also the parameters of the customer choice model. These parameters are usually not directly observable, but manifest themselves only indirectly in booking behavior over time. This adds a whole new level of complexity and difficulty to the forecasting process, especially so, when combined with the data-sparsity issues that arise in network-based revenue management.

This complexity seems to slow-down the implementation of choice-based revenue management in practice. Weatherford & Ratliff (2010) report in their literature review that choice-based revenue management promises more than 5% revenue increase, yet the adoption rate is relatively low. The share of surveyed airlines that run "low-fare competitor modules" is 22%; a category that includes full-scale choice-based revenue management systems, but also much simpler manual or rule-based approaches.

The low adoption rate despite significant potential revenue gains is evidence that there is a gap between the state of the art in the literature and what is needed to implement choice-based revenue management systems in practice. We believe that this gap is mainly to be found in the issue of demand estimation, especially from sparse booking data, and the goal of this thesis is to fill this gap.

1.2. Industries

This thesis focuses on the issues that arise in revenue management when both complementary and substitutable products exist, i.e. at the intersection of network-based and choice-based revenue management. Additionally, we assume that revenue management is quantity-based, and not price-based.

Table 1.1 summarizes the applicability of these three assumptions to a number of industries. We used the following rules to compile table 1.1: Complementary products exist whenever a customer wishes to purchase multiple resources in one bundle. Substitutable products exist when customers can chose between different prices or product categories. This is almost always the case, however in business to business relationships, there are often only few large customers with whom prices are often negotiated, such that modeling demand by a continuous function seems inappropriate. As another exception, casinos, theaters and sporting events implement strong demand segmentation such that there is mostly only one rate available to a particular customer. Finally, in many industries it is more natural to control prices directly instead of controlling the availability of a set of price points, as done in quantity-based revenue management.

Table 1.1 shows that airlines fulfill all three key assumptions and therefore provide a good environment to evaluate potential solutions to our research questions. Therefore, this thesis is based on the application of revenue management in the airline industry and our simulation models are calibrated towards that. However, other industries fulfill the key assumptions as well and, in principle, our proposed methods should be applicable to such industries as well.

1.3. Outline

In the remainder of the first part of this thesis, we review the relevant literature in chapter 2 and identify the research gap that this thesis aims to fill in chapter 3. Then, part II investigates demand estimation in the presence of customer choice, but without complementary products, such that data sparsity is not yet an issue. Before analyzing concrete demand estimation procedures in chapter 6, we introduce our model of the revenue management process in chapter 4 and set up a simulation environment in chapter

Industry	Complementary	Substitutable	Quantity-based
Airlines	+	+	+
Hotels	+	+	+
Car Rental	+	+	+
Retail	-	+	-
Advertising	o	o	-
Natural Gas Transmission	+	o	-
Electricity	-	+	-
Tour Operators	+	+	+
Casinos	+	-	+
Cruise Ships	-	+	+
Passenger Railways	+	+	+
Air Cargo & Freight	+	o	-
Theaters & Sporting Events	-	-	o
Manufacturing	+	o	-

Table 1.1.: Applicability of three key assumptions – the existence of complementary products, substitutable products and quantity-based control – in a variety of industries with an existing practice of revenue management(compiled from Talluri & Ryzin 2005, chapter 10); applicable (+), somewhat applicable (o) or not applicable (-)

5. Next, in part III of this thesis, complementary products – and alongside those the issue of data sparsity – enter the picture. Chapter 7 studies and quantifies the impact of data sparsity on estimation quality and achieved revenue, while chapter 8 presents a potential mitigation strategy. Finally, chapter 9, part IV concludes this thesis by summarizing our findings and elaborating on limiting assumptions and future research.

2. Literature Review

Littlewood (1972) is commonly regarded as the first publication on revenue management, focusing on selling a single resource at two distinct prices. In the 40 years since then, a wide body of literature has developed and we refer the reader to the text books of Talluri & Ryzin (2005) and Phillips (2005) for a general overview. Furthermore, McGill & Ryzin (1999) and Chiang et al. (2007) provide extensive reviews of the field, the former also including a glossary of revenue management terminology. Here, we present those articles from the revenue management literature that explicitly consider demand estimation or learning. Moreover, we review publications on the problem of small numbers and data sparsity in revenue management and statistics. Literature that provides the background for specific methods proposed in this thesis is deferred to the respective sections.

2.1. Revenue Management with Demand Learning

In much of the existing revenue management literature, the underlying demand model is assumed to be known and the focus is on improving the optimization methodology. This gap in the literature is – among others – acknowledged by Weatherford et al. (2003) and later re-stated by Lin (2006) and Bobb et al. (2008). However, there is a small stream of articles concerned with dynamic pricing problems that is closely related to our work in that it considers the problem of demand learning explicitly.

Bitran & Wadhwa (1996) consider a dynamic pricing problem for seasonal products. They model customer arrivals by a Poisson process with known, but potentially time-variant rate. Each customer has a reservation price that is drawn from a probability distribution with a potentially unknown parameter and may also vary over time. There is only a single product, which an arriving customer will purchase if her willingness-to-pay exceeds the price of the product. Bitran & Wadhwa develop a Bayesian update procedure for the parameter of the reservation price distribution while assuming that the arrival rate is known. They allow for some demand changes between time periods, but these changes cannot be random and have to be known to the modeler.

Lobo & Boyd (2003) also consider a dynamic pricing problem, using a linear demand model with an intercept and one coefficient corresponding to price. The parameters of the model are unknown and drawn from a Gaussian distribution. The authors provide equations of the Bayesian update, equivalent to the Kalman Filter equations. They consider the active learning problem in this setting. In traditional learning, demand estimation is passive in that it only observes prices or availabilities and has no direct influence on them. There can be an incentive however, to select prices or availabilities that facilitate better demand estimation. This will of course result in short-term revenue losses, but these may be outweighed by future revenue gains due to the improved demand estimates. Lobo and Boyd develop an approximate solution to the active learning problem in this setting, making use of convex semi-definite programming techniques. Our work on demand estimation is closely related to this, however we consider revenue management, instead of dynamic pricing and allow for additional flexibility in the specification of the demand model.

Carvalho & Puterman (2004) assume a log-linear demand function with unknown parameters and rely on the Kalman Filter equations for Bayesian learning. They develop a one-step-look-ahead strategy based on a second degree Taylor expansion of the expected future revenue as a heuristic solution to the active learning problem. The authors compare this pricing policy to various other schemes, including a myopic strategy, random price variation and a "softmax" strategy, in Monte Carlo simulations. In their setting, the myopic policy clearly underperforms compared to the other pricing policies, and the one-step-look-ahead strategy yields slightly higher expected revenues than the remaining pricing schemes. Modifying the model to consider binomial demand, such that only a single customer may arrive in each time-period, leads to qualitatively identical results (Carvalho & Puterman 2005). While Carvalho & Puterman developed this model with web traffic in mind, it can also be a reasonable assumption for very low-volume airline markets.

Aviv & Pazgal (2005) provide a model of Bayesian demand learning where customers arrive in a Poisson process with unknown rate. They model the uncertainty about the arrival rate as a Gamma distribution to achieve a simple update rule for the belief distribution. Price-sensitivity is modeled with an exponential distribution with a known mean. The authors focus on the distinction between active learning and passive learning. In their setting, they show that the benefits of active learning are minor as long as the level of uncertainty is not too high, in contrast to the results in (Carvalho & Puterman 2004). Thus, they conclude that the passive learning approach is a reasonable heuristic.

Intuitively, this result seems to stem from the fact that only the arrival rate is uncertain, but price-sensitivity is known.

Şen & Zhang (2009) also model demand as a Poisson process with unknown arrival rate and a distribution of reservation prices that is unknown, but from a finite set of candidate distributions. They provide a Bayesian learning model to estimate the arrival rate and the reservation price distribution jointly. The information requirements of their method increase with the number of candidate distributions, making it crucial for practical implementations to restrict the candidate set as much as possible.

Araman & Caldentey (2009) also model demand as a Poisson process and evaluate both linear and exponential dependence on price. Price elasticity is assumed known, but the overall arrival rate is unknown, but restricted to a finite set of potential values. Using this demand model, they consider a dynamic pricing problem, in this case however for *non-perishable* products.

Vulcano et al. (2010) consider a model where customers arrive in a Poisson process and then choose among the subset of currently available products according to a multinomial logit choice model. They use an expectation-maximization (EM) procedure to find a maximum-likelihood estimate of the demand parameters, both from simulated and from real-world data. Vulcano et al. (2012) present a re-formulation of this estimation problem in terms of so called Primary Demand that is the demand for a product in case all products were available. The re-formulation yields a much simplified EM procedure to estimate both of the arrival rate and the product valuations. The method developed here is not a demand learning method per se, however, since it does not allow for incremental updates in the demand estimates. Instead, all historical data has to be processed every time an updated demand estimate is required.

Gallego & Talebian (2012) examine a similar model, with Poisson arrivals and a multinomial logic choice model. However, here the seller is assumed to offer different "versions" of a single product with related utilities. They present a method to jointly estimate the common utility value of all versions, the "core value", and the overall arrival rate using maximum-likelihood estimation. As Vulcano et al. (2012), this procedure does not allow for incremental updates. Furthermore, there is no demand censoring, since the seller observes lost sales due to exhausted capacity, opposed to the practice in airline revenue management.

A slightly different approach in modeling the demand is taken by Stefanescu (2009) and Kwon et al. (2009). Stefanescu models demand as a multivariate Gaussian distribution. She argues that customer choice modeling may not be appropriate in the face of customer heterogeneity or missing data, e.g. about the choices offered by competitors. Time and inter-product dependence can be modeled through demand correlation in her model. The author develops an EM-algorithm to estimate this model given censored data which shows promising results. Again, this is not strictly demand learning, since the EM-algorithm requires the complete data set to update the current estimate. Moreover, the descriptive nature of this demand model seems less suited as an input for revenue optimization.

Kwon et al. (2009) consider "non-cooperative competition among revenue maximizing service providers" in a dynamic pricing context. Each firm uses a Kalman Filter to estimate the parameters of the demand model. Demand is deterministic and independent between different products, but it depends exclusively on past and current market prices in a linear fashion. The authors assume that the coefficients of their demand model evolve according to a random walk, similar to the assumption in this thesis. However, they model the dynamics of demand parameters only over a single, continuous selling horizon. While this model seems appropriate in a retail setting, it does not realistically capture demand dynamics in airline revenue management, where demand evolves both over the selling horizon of a particular flight and between consecutive flights.

Li et al. (2009) and Chung et al. (2012) extend the model of Kwon et al. (2009) by allowing for a much more general form of demand evolution over the selling horizon. Moreover, they highlight the notion of a state-space model to formulate the dynamic pricing and demand estimation problem and use a Markov chain Monte Carlo technique for parameter estimation. Yet, their demand model is still very limited, in that it does not include stochastic demand, dependence between products nor demand evolution between consecutive flights.

Similarly, Lin (2006) assumes that there is an existing, but somewhat uncertain forecast in the form of a Gamma distribution over potential customer arrival rates and focus on learning demand over the selling horizon. They propose a method to update the initial forecast in real-time and show through numerical experiments that their method can significantly improve revenue. Applicability to the airline revenue management problem is limited however by the assumption of a known, constant willingness-to-pay distribution and a known arrival rate trajectory over the selling horizon.

Besbes & Zeevi (2006) split the selling horizon in an initial learning period, in which a general, non-parametric demand function is estimated from a Poisson customer arrival process. After the initial learning period, a static price is computed from the information gathered up to that point. The authors show that this two-part strategy is asymptotically optimal when both capacity and customer arrival rate tend to infinity. Nevertheless, in our setting of small arrival rates and gradually changing demand, this strategy seems inappropriate.

While demand learning during the selling horizon may offer additional benefit in the area of airline revenue management, as noted by Lin (2006), the primary concern is to update demand estimates between consecutive flights, which is therefore the focus of this thesis. Despite of this fundamental difference, we model the demand estimation problem as a state-space model, similar in spirit to Chung et al. (2012), and an additional real-time update procedure as in Lin (2006) might be a possible extension of our work.

The thesis of Boyer (2010) is most closely related to our work. In the context of the Passenger Origin Destination Simulator (PODS), Boyer analyzes methods for estimating passenger willingness-to-pay from booking data and also proposes a clustering method to find the appropriate aggregation level for demand estimation. We extend his work in several directions. First, his estimation methods focus solely on the price-sensitivity parameter of demand and uses a two-step procedure, which first estimates parameters per time-period and then fits a curve to these initial estimates. We propose methods that produce estimates for all demand parameters simultaneously and in a single step, combining all available information to find the best overall estimate. Second, Boyer employs a standard clustering algorithm (k-Means), using the difference between his regression parameters as the distance metric. We develop the theory that allows us to assess the impact of data size on the quality of our forecasts, and use that to derive a clustering algorithm that strives to minimize forecast error explicitly.

2.2. Data Sparsity and Small Numbers in Forecasting

As Talluri & Ryzin (2005, p. 83) note, data sparsity issues enter the practice of revenue management once the move from flight-based to network-based revenue management is made. This problem has been acknowledged multiple times in the revenue management literature – for airlines and other industries (see Bartke et al. 2013). In that article, we also present data on the number of forecasted entities in a large airline network: It is orders of magnitude higher than the number of booking events and grows with increasing sophistication of the underlying revenue management model. Vulcano et al. (2010) briefly investigate the effect of two different base arrival rates on demand estimation and find that forecast quality is greatly diminished for smaller arrival rates. However, with the exception of Boyer (2010), we are unaware of explicit inquiries into how to avoid the impact of data sparsity in the revenue management literature. In this thesis, we aim to provide such an investigation in chapter 7.

In the statistics literature the related problem of small sample size is well-known. Schmid (2011) reviews methods for *small area* or *small domain* estimation. In this model, sample size per area or domain is too small to yield reliable estimates directly. To overcome this, correlations between spatially close areas are exploited to improve estimation accuracy.

The data sparsity problem in airline revenue management, however, does not stem from small sample sizes. Each available product/price-point combination that is offered by an airline for some fixed period of time constitutes an observation or sample. Data sparsity implies that most of these observations will be of zero bookings. Yet, they are still observations. In that sense, sample size is large, but observed numbers are small. Therefore, we call our problem the *problem of small numbers*. The distinction between small sample size and small observation size is discussed in more detail in chapter 7.

A similar problem also arises in other application areas where empirical data is collected. Since the collection of samples or the conduction of experiments is usually costly, the goal is to extract as much information as possible from a given set of data. Here, similar to the problem of small numbers in revenue management, the question arises how far the data set can be broken down before sampling noise starts to dominate the results. A number of authors have acknowledged this fact and proposed techniques specific to their respective fields, e.g. Roff & Bentzen (1989) for inference from contingency tables in biology, Agarwal et al. (2007) for analysis of click-stream data, or King & Zeng (2001) for analysis of rare events in the political sciences. These articles are concerned with estimation of event rates, their confidence intervals and statistical inference on those rates. Unfortunately, our problem is slightly more involved since we are not only interested in estimating arrival rates, but need to find the parameters of an underlying demand model that generates those rates. Therefore, we were unable to directly transfer any of the proposed methods to our problem.

Duncan et al. (1993) face a similar problem in the context of economic time-series forecasting. They wish to forecast a number of related time-series, e.g. from adjacent geographical areas similar to Schmid (2011), and create a hierarchical model to describe the relationship between these individual time-series. Their method lets them combine information from each observed time-series with cross-sectional information to yield more accurate forecasts. In chapter 8, we describe how our assumptions and our forecast merging method differs from that of Duncan et al..

In addition to small arrival rates, other effects may further increase the number of zero-observations. For example, inaccurate availability information may lead to zero observed bookings for products that were thought to be available for sale when in fact they were not. Ridout et al. (1998) review approaches to model such situations where the number of zero-observations is larger than the Poisson distribution would suggest. In this thesis, however, we assume that historical availability information is sufficiently accurate such that the impact of any remaining inaccuracies is negligible.

3. Research Gap

While the issue of data sparsity in airline revenue management has been acknowledged as early as 2005 in the well-known book by Talluri & Ryzin (2005), little research has been published on the details of this problem and on how to solve it. As noted above, the exception here is Boyer (2010) who proposes a standard clustering algorithm to address the problem of small numbers. However, even here, no thorough investigation into the structure and root causes of the problem of small numbers is conducted, and consequently the proposed solution cannot exploit any of that special structure of the problem. This is the research gap that this thesis primarily addresses: Proving the existence of the problem of small numbers, revealing its structure and build improved, practical solution strategies based on that.

Before the data sparsity issues can be considered, the general question of how to adequately estimate demand, independent of data sparsity, needs to be answered. As evident from section 2.1, a number of authors have proposed methods to estimate demand from booking data in their specific setting. Yet, no systematic treatment of the demand estimation problem for general, choice-based revenue management systems along with a thorough comparison of competing methods exists. While a truly comprehensive treatment in this sense is beyond the reach of a single thesis, we aim to fill this gap in the literature to the point where we can start to address the primary research question of this thesis: the problem of small numbers. At the same time, we wish to provide a foundation on which other authors can build to complete the picture.

Thus, following this introductory part, the second part of this thesis addresses the demand estimation problem in general. First, in chapter 4, we develop a new representation of the revenue management process in general and the demand estimation problem in particular. The goal of this model is to abstract away from any concrete revenue management method by assuming no specific demand model or optimization method. Such a model allows us to formulate new demand estimation procedures that are not tied to a specific demand model, and lets us examine the structure of the demand estimation problem independently of a particular revenue management method.

Then, chapter 5 introduces the simulation model that all simulation studies in this thesis are based on. The objective of this chapter is to build a test-bed for our proposed methods that produces results that can be confidently generalized to real-world settings. A strong focus is therefore on building life-like scenarios that are still abstract and general enough that the simulation results are not governed by particularities of the specific scenarios.

Concluding the part on demand estimation, chapter 6 finally addresses the demand estimation problem itself. Here, we aim to find new demand estimation procedures that significantly improve on existing methods and perform relatively close to the theoretical optimum. Using the framework developed in chapter 4, two novel demand estimation procedures are presented and evaluated in the simulation setting from chapter 5.

The third part then focuses on the overarching research question of this thesis, the problem of small numbers in demand estimation. First, in chapter 7, we pose the question whether such a problem of small numbers really exists and how to precisely define it. We answer this question in the theoretical framework from chapter 4 to yield a general statement that does not depend on any particular revenue management method and with a simulation study to assess the practical implications.

The objective of chapter 8 is then to mitigate the problem of small numbers as much as possible. The results from chapter 7 provide the insight that there is a fundamental trade-off between forecast granularity and stability, which cannot be broken by improved estimation procedures. The more specific goal of chapter 8 is therefore to find the optimal trade-off between the two such that the negative impact of the problem of small numbers in terms of forecast accuracy and achieved revenue is minimized. We propose a method to accomplish this and evaluate it in a simulation study.

Finally, in the fourth part of this thesis, we summarize our findings, provide insight on potentially limiting assumptions and set directions for future research in chapter 9. The appendix in chapter 10 contains additional technical details and charts that have been left out of the main text since they were not central to the discussion.

Figure 3.1 provides a high-level overview of the respective objectives of part II and part III of this thesis. Leaving classical revenue management methods in the top-left corner behind, part II addresses the issue of choice-based demand estimation, under the assumption that data sparsity is not an issue. Subsequently, part III lifts that assumption, and focuses on the problems that arise through data sparsity and how they can be mitigated. Since data sparsity really only becomes a serious issue under choice-

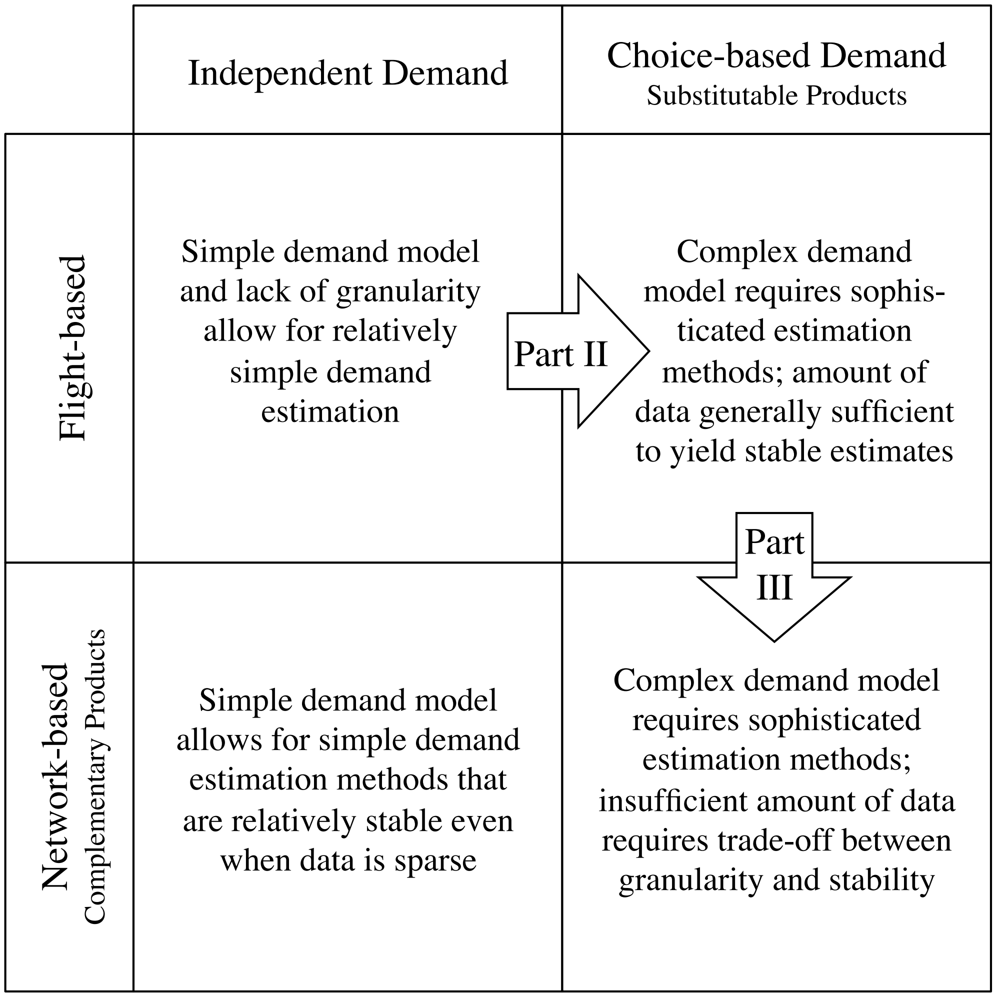

Figure 3.1.: Conceptual overview of the research objectives of the two central parts of this thesis.

based demand models[1], the path this thesis takes through figure 3.1 splits the overall problem more evenly than the path over the bottom-left corner would do.

[1] A point for which chapter 7 provides evidence

Part II.

Demand Estimation

4. State-Space Model for Demand Estimation

This chapter introduces a state-space model for demand estimation. Towards that, we first formulate a general and abstract model of the revenue management process as a whole. Then, we zoom in on the demand estimation problem, providing a consistent terminology in section 4.2, developing the actual state-space model in section 4.3 and investigating its information structure in section 4.4.

4.1. A General Model of Revenue Management

We consider an abstract revenue management system in which the estimation and forecasting module is embedded. The revenue management system is general, in the sense that it assumes no specific methods for demand estimation, forecasting, optimization or inventory control. We do assume however that there is a separation into forecaster, optimizer and inventory modules and that the system controls the availability of booking classes, instead of setting prices directly. There are forecast-less revenue management methods, such as reinforcement learning, and there is a large body of literature on dynamic pricing, both of which violate the above assumptions. In the traditional airline industry, nevertheless, the additional value of demand forecasts beyond revenue management (e.g. for fleet assignment) and the pervasive use of the booking class standard throughout the distribution, booking and check-in processes has so far prevented the adoption of these newer methods in practice (Talluri & Ryzin 2005, p. 176) and will likely do so in the foreseeable future. For that reason, we believe our framework of a revenue management system is general enough to describe the general workings of most airline revenue management systems used by traditional network carriers. Figure 4.1 provides an overview of the complete revenue management system. In the remainder of this section we will provide descriptions of the individual parts.

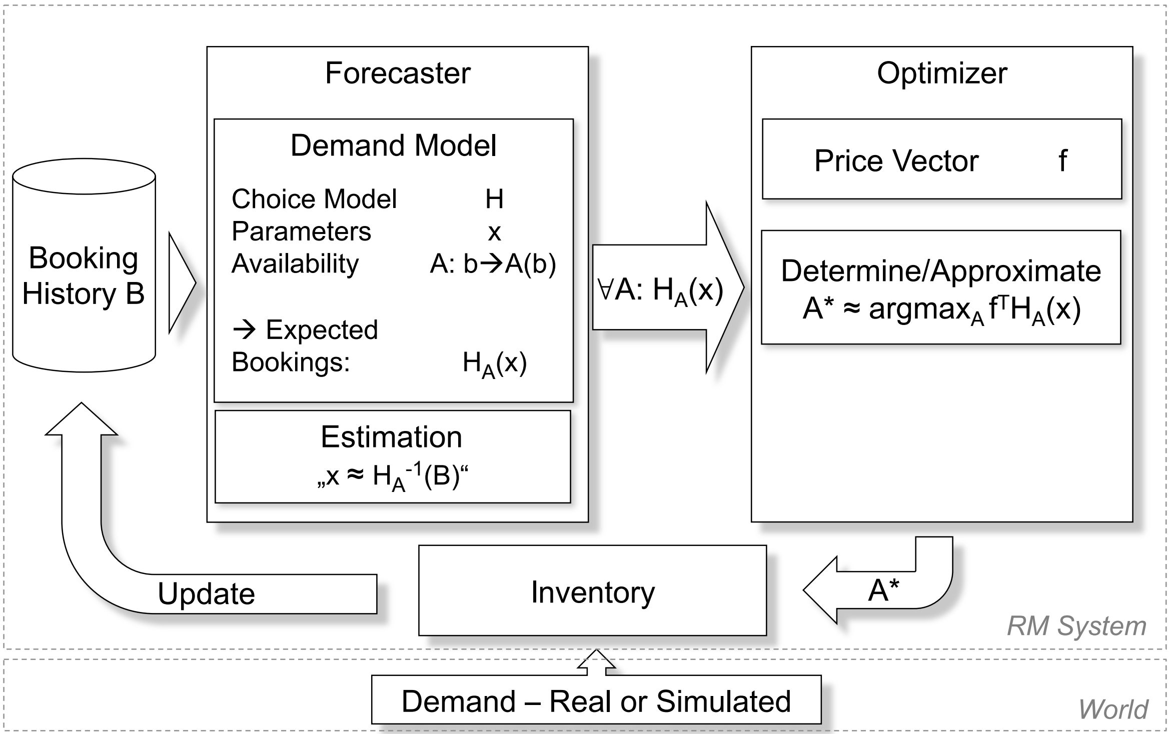

Figure 4.1.: Model of a traditional airline revenue management system

4.1.1. The Forecaster

The forecaster module is defined by a demand model and an estimation procedure for that demand model. The demand model provides a mapping H_a from product availabilities a and a vector of demand parameters x to an expected number of bookings for each product $H_a(x)$. Here, a product is any individually salable offering of the airline, typically a specific booking class for a specific itinerary through the airline's network on a particular date and time, sold a specific number of days before departure in a particular country. The availability of a certain product will usually depend on the number of seats already sold for that product or on the flights involved in the itinerary. Hence, we express the availability as a function of the current vector of bookings b.

The choice model $H_a(x)$ can be any function of a and x, as long as the restrictions on permitted sales for each product as prescribed by a are obeyed. E.g., $H_a(x)$ could assume independent demand, in which the demand for a certain product only depends on the availability of the product itself. In the most general case, demand for some product could depend on the availabilities of all other products. The former is not a very realistic assumption but has been used in practice due to its simplicity. The complete model provides the most flexible description of demand, but with billions of offered products for a large airline, this model quickly becomes intractable. Realistic demand models will therefore severely restrict the set of dependencies, e.g. to the set of booking classes for the same itinerary or to a set of similar itineraries for the same origin and destination pair.

The parameter vector x encapsulates whatever is a priori unknown about the demand model. The length of the vector and the interpretation of its components depends on the concrete demand model. In an independent demand model, x may simply represent the expected bookings per product if said product is available. For dependent demand models, at least some components of x have to describe the dependency of demand for a certain product on other products availabilities. This may be implemented, for instance, by including a price-sensitivity parameter in x or by modeling demand overlaps, that is demand that will realize in the cheapest available product from some set of products. Additionally, x may also contain seasonality factors or weekday patterns.

Since x is unknown to the modeler, it has to be estimated from observed sales data. That is, we try to find the probability distribution of x, conditional on the history of booking vectors $B = \{b_1, \ldots, b_T\}$ and the history of availabilities $A = \{a_1, \ldots, a_T\}$. Alternatively, we might only obtain some property of this distribution, such as its mode which would yield the maximum-likelihood estimate, or its mean which would yield the expected value of x conditional on the observations. We defer further discussion of the estimation problem to sections 4.2–4.4 and continue with the assumption that we have some estimate of x, either in form of a probability distribution or a point-estimate.

Finally, the forecaster will compute the expected bookings (or the distribution thereof) for all feasible availability functions using the current demand estimate x. Availability functions are feasible if they obey the capacity constraints and are implementable by the inventory. The predictive power of the demand model is crucial for this step, since it allows the forecaster to provide the number of expected bookings even for availability situations which were never observed in the past. In practice, no forecaster simply iterates over all feasible availability functions and computes a list of expected booking vectors. Such a list would be prohibitively long. However, the structure of concrete demand models such as limited dependence between products usually allows for a much more efficient encoding of this information. Moreover, the optimization method used may also limit the amount of information that is relevant for the optimization step. Information that isn't used by the optimizer need obviously not be computed by the forecaster in practical implementations.

4.1.2. The Optimizer

The optimizer combines the expected bookings $H_a(x)$ for all a received from the forecaster with a vector of prices of each product f. It then finds the availability a^* that maximizes $f^T H_{a^*}(x)$. This conceptually simple step can usually not be solved to optimality in practice due to the very large number of potential availability functions a. Hence, in actual implementations heuristic methods are used to find an approximate solution to the optimization problem.

The output of the optimizer is the optimal availability function a^* which is sent to the inventory.

4.1.3. The Inventory

The inventory implements the availability function a^*: It offers all products for which $a^*(b)$ is true and prevents the sale of all other products, b being the current vector of booking on hand. Actual inventory systems will constrain the set of availability functions that can be implemented. In practice, inventory systems may constrain the number of bookings in a certain booking class on a particular flight by booking limits or protection limits. Alternatively, they can set a bid-price for each flight and only make those products available for which the price exceeds the sum of bid-prices of the flights in the itinerary. Combinations are also possible and in practical use.

After some pre-defined time-period, the booking history B and the availability history A are updated and augmented by the new observations. This will then trigger a new loop through the complete revenue management system, from estimation to prediction to optimization to implementation.

4.1.4. The Feedback Loop

The fact that revenue management systems have a feedback loop renders their long-term, dynamic behavior non-trivial. Even for the case of relatively simple constituting components the overall behavior may be complicated and hard to predict. Choosing an availability now will influence the observations in the upcoming time-period which will influence the new demand estimate. This may in turn lead to a revised availability. Specifically, if something akin to price-sensitivity is to be estimated, it may well be worth it to occasionally set availabilities that are not short-term optimal in order to learn more about this price-sensitivity. If instead only the short-term optimal availability would be used at all times, no new information about the price-sensitivity parameter is gained. In this case, the system might get stuck in a state far away from the actual optimum.

In control theory, the problem of simultaneously optimizing an objective function now and improving knowledge about process parameters for future optimization is known as dual control. For linear systems with additive Gaussian noise and quadratic objective functions, the certainty-equivalence principle holds (Water & Willems 1981). It states that despite the feedback loop, optimization and estimation can be treated independently, since the optimum is independent of forecast uncertainty. In our problem, the optimum is the solution of a complex optimization problem, and as such is not a simple quadratic function of the demand estimate. Therefore, the certainty-equivalence principle cannot be expected to hold in our case.

Easley & Kiefer (1988) analyze the general problem in which a decision maker's action influence both an immediate reward as well as future knowledge about an unknown parameter of the reward function. They focus on the asymptotic behavior of this learning problem, and show that the decision maker may be content with incomplete knowledge of the demand parameter. However, in a realistic revenue management system the analysis of the feedback loop may be close to impossible. As Wittenmark (1995) notes, finding the optimal dual control is very difficult, even for relatively simple problems, and we are aware of only three articles (Lobo & Boyd 2003; Carvalho & Puterman 2004; Aviv & Pazgal 2005) that consider the feedback loop explicitly in a pricing or revenue management context. In our work, we evaluate the option of using a simple "active estimation" heuristic that randomly perturbs the estimates when uncertainty is high. Results are reported in section 6.5.1.

4.1.5. Choice Functions

While the general discussion remains independent of a particular choice function, it is illustrative to consider a few examples. Moreover, the simulation study has to use a particular choice function, which will also be introduced here.

Independent Choice The simplest conceivable choice function is the independent choice function. Here, demand for each of the airline's products is independent from demand for the airline's other products. As a consequence, expected bookings for a single product i only depend on the availability of that particular product i: if it is available, we expect x_i bookings to occur and otherwise none. This can be written as the product between a diagonal "availability matrix" and the parameter vector x:

$$H_a(x) = diag(a) \cdot x \tag{4.1}$$

where $diag(a)$ is the matrix with the vector a as its diagonal and zeroes everywhere else.

Assume, e.g. there are three products sold by the airline, represented by the booking classes A, B and C. If the airline made all booking classes available for sale, they'd expect 1, 3 and 5 bookings in those classes, respectively; i.e. $x = (1, 3, 5)^T$. Now assume that only A and B were available. Then $a = (1, 1, 0)^T$ and

$$H_a(x) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix}. \tag{4.2}$$

Independent choice functions were and are widely used in the practice of revenue management. E.g. the first widely deployed revenue management algorithm, EMSR (Belobaba 1989), relies on the independent demand assumption. As argued in section 1, this was justifiable by the relatively strong fencing mechanisms used by traditional carriers in the past. However, the more flexible fare structures that are being introduced currently, make the independent demand assumption much more problematic.

Market-sensitive Choice Winter (2010, 2012) presents an approach to handle so-called buy-down behavior. The demand parameter vector x has two parts. First, it gives the expected bookings for each class, in the case that this class was the only available one. These numbers are called "attainable demand". Additionally, there is a buy-down graph which has a node for every booking class and directed edges between those classes where buy-down can occur. The set of class-pairs between which buy-down can occur is determined from prices and restrictions of the respective classes, the exact derivation, however, is quite involved and beyond the scope of this section. The important point for our purposes is, that the second part of x contains a buy-down value for each edge in the buy-down graph. This buy-down will realize if the destination class of the buy-down edge is available. In that case the buy-down demand is subtracted from the attainable demand of the originating class to yield the final number of expected bookings.

This choice function can still be written as a matrix product, i.e. it is a linear function in x. However, here the matrix is no longer diagonal, but may contain negative ones off the diagonal to encode realized buy-down.

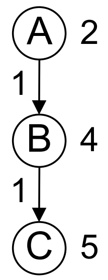

Figure 4.2.: Example of a buy-down graph for three booking classes with attainable demand and buy-down numbers

Assume again, that there are three booking classes A, B and C. Figure 4.2 shows a potential buy-down graph for this scenario, along with attainable and buy-down numbers. The demand parameter vector would then be $x = (2, 4, 5, 1, 1)^T$. If only class A and B were available, i.e. $a = (1, 1, 0)^T$, then

$$H_a(x) = \begin{pmatrix} 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 4 \\ 5 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \\ 0 \end{pmatrix}. \tag{4.3}$$

The advantage of this choice function is that it can explain any choice behavior given an appropriate buy-down graph. In contrast to other approaches, it does not aim to explain customer behavior from latent variables, such as price-sensitivity and (dis-)utility for product restrictions. In that sense, market-sensitive forecasting is a non-parametric approach. However, the afforded flexibility is also a disadvantage. First, the demand parameter vector depends on the buy-down graph and through that on product prices and restrictions. However, these change frequently in the airline industry, and with every such change the parameter vector has to be either discarded entirely or adapted to the new buy-down graph. Second, there is no clear distinction between parameters that describe demand volume and parameters that describe customer choice. In chapter 7 we will show that there is a fundamental difference between how those two parameter sets behave in the face of data sparsity. Thus, a separate treatment of these parameter sets is appropriate in that situation. This, however, is only possible when the choice model affords the separation into volume and choice parameters.

Price-sensitive Choice A large number of choice functions can be constructed by combining a demand volume parameter x_{vol} with a discrete choice model $C_a(x_{choice})$. The choice function then has the form

$$H_a(x) = x_{vol} \cdot C_a(x_{choice}). \tag{4.4}$$

A particularly simple example of this type of choice function is the price-sensitive choice function. Here, the only choice parameter is a price-elasticity parameter. Customers are assumed to ignore all booking class restrictions and thus always purchase the cheapest available product or not purchase at all. Hence, $C_a(x_{choice})$ is

$$C_a(x_{elast}) = \begin{cases} \exp\left(-x_{elast}(\frac{f_i}{f_{base}} - 1)\right) & i \text{ is cheapest available product in } a, \\ 0 & \text{otherwise.} \end{cases} \tag{4.5}$$

where f_i is the price of product i and f_{base} is the price of the cheapest product overall. This choice function is also known under the name of Q-forecasting, described in detail by Cléaz-Savoyen (2005), and has also been used by Gallego & Ryzin (1994, 1997).

Again, assume there are three classes A, B and C with only the first two being available. Further let the prices be 300, 200 and 100, $x_{vol} = 5$ and $x_{elast} = 1$. Then,

$$H_a(x) = 5 \cdot C_a(1) = 5 \cdot \begin{pmatrix} 0 \\ \exp\left(-1 \cdot (\frac{200}{100} - 1)\right) \\ 0 \end{pmatrix} \approx \begin{pmatrix} 0 \\ 1.84 \\ 0 \end{pmatrix}. \tag{4.6}$$

The appeal of this choice function is its simplicity. Customer behavior is explained by just two parameters, demand volume and price-elasticity. If an airline uses no or very few booking class restrictions, this simplistic model may yet be justified.

Hybrid Choice In practice, an additive combination of a price-sensitive and an independent choice function is used under the name of Hybrid Demand (Walczak et al. 2009) or Hybrid Forecasting (Reyes 2006).

Formally, the Hybrid choice function has the form

$$H_a(x) = x_{vol} \cdot C_a(x_{choice}) + diag(a) \cdot x_{product}. \tag{4.7}$$

where $C_a(x_{choice})$ is the price-sensitive choice function, defined in equation 4.5. Since it is both simple and of practical relevance, we use the hybrid choice function in the simulation studies throughout this thesis.

4.2. Demand Estimation, Forecasting and Expected Bookings

In the very early literature on revenue management little distinction was made between expected bookings, demand estimates and forecasts. This was appropriate since less complexity in the demand models allowed for this direct identification. For the purpose of this article however, a clear distinction will be useful:

- Demand Estimate: The current belief about the parameters of the demand model, given all past observations. This can be represented as a probability distribution over the parameter space of the demand model, conditional on all past observations: $p(x_t|B_t, A_t)$. Finding this distribution is the problem of demand estimation. With an independent demand assumption, there is a simple relationship that maps the parameters x_t of the demand model and the availabilities A_t to the observed bookings B_t. In this case, demand estimation is straight-forward. When the relationship becomes more complex, more sophisticated statistical methods will be required.

- Forecast: Given the current belief about the parameters of the demand model at time t, what is the belief about the parameters at some time $t' > t$? In the simplest case, we would assume that there are no systematic demand changes over time, such that the forecast equals the demand estimate, that is $p(x_{t+k}|B_t, A_t) = p(x_t|B_t, A_t)$ for any $k \geq 0$. In an industry with strong seasonal effects and weekly demand patterns, this will rarely be appropriate. So in general, the former equation will not be true. Instead the probability distribution of the demand model parameters at some future point in time, conditional on the current belief $p(x_{t+k}|p(x_t)) \neq p(x_t)$ has to be considered. In this thesis, however, we ignore seasonal or other such effects, and therefore demand estimates and forecast will be identical. Thus, we also use both terms interchangeably.

- Expected Bookings: For revenue optimization, we will essentially be interested in the expected bookings in some future period for any given availability situation, such that we can find the revenue optimizing one. The demand model chosen will prescribe the relationship between the parameters of the demand model and the expected bookings. Since the parameters are not known and we only have distributional information about them, there will also be uncertainty about the expected bookings. This can be expressed as a probability distribution conditional on the parameter distribution: $p(E[B_{t+k}]|p(x_t))$.

The act of estimating and forecasting demand can be interpreted in a Bayesian framework. When adding a new observation to the booking and availability histories, the forecast distribution $p(x_{t+1}|B_t, A_t)$ acts as a prior distribution for the Bayesian update. Similarly, the one-step forecast $p(x_{t+1}|B_t, A_t)$ can be derived from the current demand estimate $p(x_t|B_t, A_t)$ and the time-evolution distribution $p(x_{t+1}|x_t)$ using Bayes' rule. The new, posterior belief $p(x_{t+1}|B_{t+1}, A_{t+1})$ can thus be found as

$$\begin{aligned} p(x_{t+1}|B_{t+1}, A_{t+1}) &\propto p(b_{t+1}, a_{t+1}|x_{t+1}) \cdot p(x_{t+1}|B_t, A_t) \\ &\propto p(b_{t+1}, a_{t+1}|x_{t+1}) \cdot p(x_{t+1}|x_t) \cdot p(x_t|B_t, A_t) \end{aligned} \tag{4.8}$$

In the following time step, this posterior distribution will act as the new prior distribution.

Equation 4.8 shows that it is not necessary to keep track of the booking and availability histories B_t and A_t, as long as the belief distribution $p(x_t|B_t, A_t)$ is known. In that sense, all available historic information is compressed into the belief distribution. At each time-step, only the new observations (b_{t+1}, a_{t+1}) and the current belief distribution need to be processed to compute the new demand estimate. This is an attractive feature with regard to practical applications: real-world observation histories may contain vast amounts of data and it quickly becomes infeasible to process all of that data whenever demand estimates are updated. Both of the estimation methods we propose in chapter 6 make use of the Bayesian update rule. They primarily differ in the way that the belief distributions are approximated, since these have no closed analytical form in general.

4.3. State-Space Model for Demand Estimation

The idea of a state-space model is to describe the dynamics of an observed quantity b_t with the help of an unobserved, hidden quantity x_t. The hidden quantity x_t is known as the *state* of the system. Both the observed quantity b_t and the state x_t can be vector-valued. The state-space model defines how the state evolves over time, i.e. the map that takes x_t to x_{t+1}, and how the state manifests itself in the observed quantity b_t, i.e. the map that takes x_t to b_t. Both maps need not be independent of the time parameter t and it is common to include stochastic terms in those maps. Stochastic terms in the state evolution map capture non-systematic changes in the state, unknown to the modeler a priori, while stochastic terms in the observation function represent non-systematic imperfections in the observation or measurement process.

As argued by Åström & Murray (2008, chapter 2), state-space models were developed in the context of control theory in the late 1950s. In these models, dynamic models, previously known from mechanics, were merged with input/output models, known from electrical engineering. Additionally, error terms were added to model disturbances and inaccuracies in the model. One initial application was space flight in which a space craft has to be precisely controlled given imprecise and indirect measurements of its position, orientation, velocity etc. State-space models have been used in the context of dynamic pricing by Kwon et al. (2009) and in a related paper by Chung et al. (2012). Our model differs from that of Kwon et al. (2009) and Chung et al. (2012) by allowing for a very general demand function and for stochastic demand. Moreover, our state-space model captures demand dynamics between consecutive flights, while demand over the selling horizon is modeled using the demand function – as appropriate in an airline revenue management setting. In contrast, Kwon et al. (2009) and Chung et al. (2012) describe demand learning exclusively within the selling horizon, a model that is more suited for retail revenue management.

A state-space representation is useful whenever it leads to a simpler description of the system than directly capturing the dynamics of the measured values. In our context, a direct description would mean to directly describe the relationship between subsequent booking results. This is difficult, however, since observed bookings depend on the underlying demand, on booking class availability and are subject to stochastic fluctuations. A state-space model on the other hand yields a much more natural description by representing these effects individually. We identify the hidden state x_t with the underlying and unobserved market demand and the observed quantity b_t with the observed number of bookings. The dynamics of demand itself are then captured in the state evolution equation, while the influence of different availabilities and stochastic fluctuation in booking numbers are represented in the observation function.

In general, state-space models can be time-continuous or time-discrete. Depending on the problem domain, either form can be more natural. In our case, we opt for the time-discrete variant. While it would be natural to assume that demand evolves continuously over time, we only observe bookings at discrete points in time (e.g. once for each offered flight). Hence, demand is only relevant at these same discrete points in time and thus a time-discrete model is sufficient.

In the demand estimation problem, we assume that the real parameters of the demand model are the hidden state x and are the realizations of an AR(1)-process:

$$x_{t+1} = x_t + w_t \quad w_t \sim N(0, Q) \tag{4.9}$$

In other words, the change of demand parameters from time t to time $t+1$ is described by a multi-variate Gaussian random variable w_t with zero mean and covariance matrix Q. This is one of the simplest state dynamics possible, since there is no systematic change over time and a simple form of random disturbance. Yet it captures the essential problem that demand changes unpredictably over time, such that historic information on demand gradually looses its value. More elaborate models might include systematic demand changes, such as known trends or seasonalities, but this is referred to future research.

Some of the values in x_t might have to be constrained to some range of valid values to be meaningful inputs for the demand model. In that case we would have to assume a truncated normal distribution for w_t. We will assume in the following discussion that the real parameter values are far enough away from the bounds of the valid range such that a truncation of the normal distribution would only have a minor effect and can thus be ignored.

To complete the state-space representation, we now add the observation equation. As noted earlier, this relationship includes both the dependency of expected bookings on availabilities and underlying demand, as well as the random fluctuations in booking numbers. The demand model H_a is exactly the function that maps the demand parameters x to expected bookings $E[b_t]$ for each offered product and for a given availability a. As such, since the actual availability may differ between different time-steps, the observation equation is time-independent in our setting.

The second part of the observation equation describes the mapping from expected bookings $E[b_t]$ to (the distribution of) actually observed booking numbers b_t. Here, we assume that bookings, conditional on their mean, follow a Poisson distribution with mean $E[b_t]$. Poisson distributions are commonly used to model customer arrivals, e.g. in queuing theory (Kleinrock 1975). The underlying assumption is that there is a very large pool of potential customers, where each customer decides to book with constant probability in any given period of time. This implies that customers arrive independently of each other, meaning that a customer arrival now does not make it any more or less likely that a customer will arrive in a subsequent time-period. As in many other settings, this is a natural assumption for our case.

However, Walczak (2006) argues that it could be beneficial to allow the fact that some customer arrivals lead to more than a single booking, e.g. when a customer books multiple seats for his family. As noted by Walczak (2006), the resulting compound Poisson distribution would lead to an increased variance of bookings, but it does not add anything that's structurally new to our model. Therefore, in the interest of simplicity, we restrict our investigation to the regular Poisson distribution, leaving the extension to a compound Poisson distribution for future research.

The booking distribution might have to be censored if the availability function a only allows for a finite number of available seats. Since usually many products compete for the same seats and a majority of products requires more than one seat, this censoring may link the booking distribution of almost all products of an airline in a non-trivial way. This is the case, even if the demand model itself has no such dependencies.

We use the following heuristic to approximate the situation: We assume that we know the fraction s_i of the observation period that product i was available and the arrival rate λ_i for product i for the complete time period. Then, we let the bookings for product i be Poisson distributed with an arrival rate of $s_i \cdot \lambda_i$.

Summarizing the above, we have the following state-space model:

$$x_{t+1} = x_t + w_t \qquad w_t \sim N(0, Q) \tag{4.10}$$

$$b_t \sim Poi(H_a(x_t)) \tag{4.11}$$

Here, equation 4.10 describes the evolution of demand over time, or – in terms of a state-space model – the evolution of the system state. Equation 4.11 is the measurement or observation equation, it describes the dependence between the state x_t of the system and the corresponding observation b_t. As such, the state-space model combines the demand model H_a, time evolution equation 4.9 and the assumption of Poisson-distributed bookings.

Alternatively, the Poisson distribution in equation 4.11 can be approximated by additive Gaussian noise, which yields a slightly modified state-space model:

$$x_{t+1} = x_t + w_t \qquad w_t \sim N(0, Q) \tag{4.12}$$

$$b_t = H_a(x_t) + v_t \qquad v_t \sim N(0, R) \tag{4.13}$$

The covariance matrix R can be chosen, such that the measurement variance is identical to that of the Poisson model. However, bookings are no longer modeled as being integer, but could now theoretically take any value, even negative ones, which certainly seems counter-intuitive. Nevertheless, demand has been modeled successfully as such in the revenue management literature (see e.g. Belobaba 1989), and only the simulation experiments will be able to reveal whether this assumption is workable in practice.

Some authors might prefer to include an exponentiation operation into the observation equation to avoid the problem of negative arrival rates, i.e. $b_t = \exp H_a(x_t) + v_t$. While this might be conceptually cleaner, we believe that the additional non-linearity introduced here causes more problems during demand estimation than it solves. It would also eliminate the direct correspondence between demand parameters and expected bookings that some of the simpler demand models afford. As mentioned earlier, we largely ignore the issue of negative expected bookings during estimation. Only if the final demand estimate used for optimization falls outside of the permissible range will those values be adjusted until they are back in the valid region.

Each of these two variants of the state-space model, the original one with Poisson bookings and the modified one with additive Gaussian noise, will result in one demand estimation procedure in chapter 6. Before moving on, however, we inquire into the information dynamics of this state-space model, independently of any specific demand estimation procedure.

4.4. The Posterior Cramér-Rao Bound and the Information Matrix

One question to ask of a state-space model is whether the observations provide enough information such that the hidden state can be estimated reliably. A way to answer such a question is to analyze the Fisher information matrix of the system, or its inverse the Posterior Cramér-Rao Bound (PCRB).

The Cramér-Rao bound provides a lower bound on the mean squared error of an estimate which has to hold for any concrete estimation method. The original Cramér-Rao bound is based on time-invariant models, however there exists an extension, the Posterior Cramér-Rao bound (PCRB), which is applicable in the context of this paper. The dynamics of this bound over time for the discrete-time nonlinear filter problem were derived by Tichavsky et al. (1998). This lower bound can be computed for many models in which an exact solution to the estimation problem is not available and can thus serve as an absolute benchmark to compare approximate estimation methods against.

A closely related concept is the Fisher information matrix (Fisher 1925) which is the inverse of the PCRB.

Formally, let $g(B_t)$ be some estimator of the demand parameters x_t operating on the booking history up to time t: $B_t = \{b_1, \dots, b_t\}$. Then, under mild regularity conditions,

$$MSE = E[(g(B_t) - x_t)(g(B_t) - x_t)^T] \geq I_t^{-1} \tag{4.14}$$

where "$\geq$" means that the difference between the matrices is a positive semi-definite matrix. I_t is the Fisher information matrix and evolves according to

$$I_{t+1} = M_{t+1} + (I_t^{-1} + Q)^{-1} \tag{4.15}$$

or, equivalently [1],

$$I_{t+1} = M_{t+1} + Q^{-1} - Q^{-1}(I_t + Q^{-1})^{-1}Q^{-1}, \tag{4.16}$$

where I_{t-1} is the Fisher information matrix from the last time step, Q is the covariance matrix from equation 4.9 and M_t is the Fisher information matrix of the observation at time t. The measurement information M_t is defined as

$$M_t = -E[\Delta_{x_t}^{x_t} \log p_{a_t}(b_t|x_t)] \tag{4.17}$$

where $\Delta_y^x = (\nabla_x)(\nabla_y)^T = (\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n})^T(\frac{\partial}{\partial y_1}, \dots, \frac{\partial}{\partial y_n})$. Here, we assume that the real x_t are fixed, but unknown parameters. The expectation in equation 4.17 is therefore taken over z_t conditioned on x_t.

Using the Poisson distribution, as discussed above, the likelihood $p_{a_t}(b_t|x_t)$ becomes[2]

$$p_{a_t}(b_t|x_t) = \prod_{i:h_{a_t,i}(x)>0} \frac{(h_{a_t,i}(x))^{b_t}}{b_t!} \cdot e^{-h_{a_t,i}(x)} \tag{4.18}$$

[1] To see the equivalence, note that $(A^{-1} + B)^{-1} = (1 + AB)^{-1}A = B^{-1}(B^{-1} + A)^{-1}A = B^{-1}(B^{-1} + A)^{-1}(B^{-1} + A - B^{-1}) = B^{-1} - B^{-1}(B^{-1} + A)^{-1}B^{-1}$, provided that all the inverses exist.

[2] We only include terms with $h_{a_t,i}(x) > 0$ in the product, that is with arrival rates > 0. If the arrival rate $h_{a_t,i}(x)$ is zero, the only possible realization is $b_{t,i} = 0$ with likelihood 1. These terms can therefore be ignored.

which yields the following measurement information matrix

$$M_t = E\left[\sum_{i:h_{a_t,i}(x)>0} \frac{(\nabla_{x_t} h_{a_t,i}(x_t))(\nabla_{x_t} h_{a_t,i}(x_t))^T}{h_{a_t,i}(x_t)}\right]. \tag{4.19}$$

Equation 4.16 is a direct specialization of the equation given by Tichavsky et al. (1998). It is efficient for computing the Fisher information matrix since it requires only one matrix inversion per iteration (Q is constant and can be inverted once at the start of the iteration). For the case of constant M_t, it is also useful to compute the steady-state of the information matrix evolution, that is when $I_{t+1} = I_t$. The steady-state equation

$$X = M_t + Q^{-1} - Q^{-1}(X + Q^{-1})^{-1}Q^{-1}, \tag{4.20}$$

is a discrete, algebraic Riccati equation and there are efficient numerical methods to solve such equations (see Laub 1979). We will use this equation to find an approximate steady-state to initialize our forecaster, see section 6.4.2.

Equation 4.15 is computationally less attractive, but much simpler to interpret intuitively: We compute the minimum variance of the last time step I_t^{-1}, add the additional noise introduced by the time-evolution of x_t with the covariance Q, convert that back to an information matrix by inversion and then add the information gained by observing at time $t+1$.

Being a strict lower bound on forecast error, the PCRB provides us with a way to measure forecast accuracy in absolute terms, and not just relatively by comparing different methods. An estimation procedure with a mean squared forecast error equal to the PCRB necessarily has the lowest forecast error among all estimation methods. Such an estimation procedure is called *efficient*. In many situations however, an efficient estimator is not known. Still, heuristic estimation methods can be gauged by measuring how much their mean squared estimation error exceeds the PCRB. We define this efficiency measure as:

$$E = \frac{tr(PCRB)}{tr(MSE)} \tag{4.21}$$

where $tr(M)$ is the trace, i.e. the sum of diagonal elements, of matrix M. Since $PCRB$ and MSE both have the form of covariance matrices, their diagonal elements are non-negative. Therefore, $E \geq 0$. Moreover, since $MSE \geq PCRB$, no diagonal element of MSE can be smaller than the corresponding diagonal element of $PCRB$. Hence, $E \leq 1$.

The efficiency measure allows us to decompose the overall mean squared estimation error into the inherent difficulty of the estimation problem – captured by the PCRB – and the performance of the estimation procedure itself – captured by forecast efficiency. Without the PCRB, we would not be able to distinguish if a high forecast error is the result of an unsuitable estimation procedure or the fundamental difficulty of estimating the demand parameters.

5. Data-Driven Simulation Setup

Theoretical performance guarantees for estimation methods can usually only be given for restrictive classes of problems and only if the assumptions underlying the estimator are actually true for the data at hand. In general, as Elliott & Timmermann (2008) argue, a good forecast is not a value in itself. Instead, it generates value by leading to better decisions – in this case better revenue optimization results, and thus ultimately more revenue. Moreover, Elliott & Timmermann note, that the forecast that models the true causal relationships most extensively and accurately is not necessarily the best forecast in the above sense. Therefore, expert judgement on the forecasting model by itself is not a sufficient predictor of its real-world performance.

While testing new estimation methods in live airline revenue management systems arguably provides the most reliable performance indicator, these live tests are usually costly to implement and carry a high risk of foregone revenue. This is especially true when evaluating forecasting methods, where the effectiveness of the demand estimation methodology will usually only manifest itself after a prolonged learning period. Moreover, changing demand conditions in the real world will overlay with the effects of a new estimation method. In a simulation environment, on the other hand, conditions can be exactly replicated. Cleophas et al. (2009) argue in more detail for the use of simulation studies to evaluate forecasting methods.

We therefore resort to a simulation study to compare and evaluate our demand estimation approaches, in line with a majority of authors in the field of revenue management (Frank et al. 2008). To maximize the likelihood that our simulation results carry over to the real world, we carefully calibrate our simulation scenarios using real world data. In this section, we describe this calibration and the general experiment design.

5.1. Simulation Environment

As part of an ongoing research cooperation with Lufthansa German Airlines, we had access to the REMATE simulation environment developed at Lufthansa for research, training and decision support purposes, see Cleophas (2009), Zimmermann et al. (2011) and Gerlach (2013). REMATE allows for highly flexible scenario definitions, such that real-world scenarios can be easily modeled. As part of this thesis, we have extended REMATE with a new demand model, our proposed estimation and forecast merging algorithms, and the capability to compute the PCRB.

5.2. Simulation Scenarios

There are three base scenarios, representing domestic, continental and intercontinental markets, respectively. The number of fare classes and their prices are taken from real world Lufthansa data for exemplary markets. To simplify the subsequent analysis of the results, we restrict ourselves to a single compartment on each flight, with capacity of 100 for domestic and continental flights, and a capacity of 200 for intercontinental flights. Furthermore, there are neither cancellations nor no-shows.

A dynamic programming (DP) approach with fare transformation is used for availability optimization. Fare transformation transforms a dependent demand model into a, for optimization purposed equivalent, independent demand model. Any standard, independent demand optimization procedure can then be used to find optimal availabilities. The dynamic programming approach explicitly models the stochastic and time-dependent nature of demand and produces a so-called bid price, which depends on the time remaining before departure and the number of seats still available. The bid price is the minimum price for which the next seat is sold, so all fare classes with a lower price will not be available while all fare classes with a higher price will be. See Talluri & Ryzin (2005) for details on the DP in general, and Fiig et al. (2009) for a description of fare transformation.

For each base scenario, there are high demand, medium demand and low demand variants. In the latter, the capacity restriction is mostly irrelevant, such that bid prices are zero and optimization is purely focused on exploiting price-sensitivity. In high demand variants, on the other hand, bid prices are positive and optimization has to exploit price-sensitivity while being constrained by limited capacity. Each simulation runs for 100 departures, to give the estimation algorithms enough time to settle into a stable state. Each simulation is repeated 10 times, such that the results are averages over 10 independent demand realizations.

5.3. Customer Choice

Customer Choice behavior is modeled using the Hybrid demand model from section 4.1.5. As customary in the airline industry, we split the one-year period before a flight's departure, the booking horizon, into a set of discrete time periods, in this case 22. These time periods are not equally spaced throughout the booking horizon, but instead are scaled such that expected bookings in each one are of the same order of magnitude. For example, the first time period is 178 days long, while the last one consists of a single day only.

Demand is expected to change significantly over the booking horizon as the customer mix tends to include more and more business travelers closer to departure. Thus, we estimate an almost completely separate set of parameters for each of the 22 time periods. We have completely separate product demand and price-sensitive demand volume parameters, since the values of those parameters heavily depend on the length of a time period. It is therefore not appropriate to model these as a smooth function in time.

In contrast, the price-sensitivity parameter does not scale with the length of a time-period and can therefore be expected to change smoothly over time. We model the price elasticity parameter x_{elast} as a degree-two Lagrange polynomial in the square root of the number of days before departure. This function has three parameters, price elasticity at the beginning of the bookings horizon $x_{elast360}$, 60 days before departure $x_{elast60}$ and at departure x_{elast0}. [1] At any number of days d before departure, the price elasticity is the following linear combination of the three parameters

$$x_{elast}(d) = s_{360}(d) \cdot x_{elast360} + s_{60}(d) \cdot x_{elast60} + s_0(d) \cdot x_{elast0} \tag{5.1}$$

$$s_{360}(d) = \frac{\sqrt{d} - \sqrt{0}}{\sqrt{360} - \sqrt{0}} \cdot \frac{\sqrt{d} - \sqrt{60}}{\sqrt{360} - \sqrt{60}} \tag{5.2}$$

$$s_{60}(d) = \frac{\sqrt{d} - \sqrt{0}}{\sqrt{60} - \sqrt{0}} \cdot \frac{\sqrt{d} - \sqrt{360}}{\sqrt{60} - \sqrt{360}} \tag{5.3}$$

$$s_0(d) = \frac{\sqrt{d} - \sqrt{60}}{\sqrt{0} - \sqrt{60}} \cdot \frac{\sqrt{d} - \sqrt{360}}{\sqrt{0} - \sqrt{360}} \tag{5.4}$$

This functional form has been chosen, since it has just enough degrees of freedom to fit the empirical data well, which we present in section 5.4.3.

In total, the complete demand parameter vector has 333 entries in domestic and

[1] The choice of 360, 60 and 0 days as "anchor points" is arbitrary, approximately equal spacing on the $\sqrt{t}$ axis improves numerical stability however.

continental scenarios: for each of the 22 time periods, there are 15 parameters; product demands for each of the 14 booking classes, and one price-sensitive demand volume parameter; additionally, there are 3 parameters to describe the trajectory of the price-elasticity parameter. In intercontinental scenarios, there are only 12 booking classes, yielding a demand parameter vector with 289 entries.

Equation 4.5, which is also part of the Hybrid demand model, contains a so far undefined reference price parameter f_{base}. Analytically, the choice of the reference price is completely arbitrary, since the demand parameters x_{volume} and x_{elast} can be adapted to recover the identical choice behavior for any reference price. Setting the reference price to the lowest existing price ensures that $\exp\left(-x_{elast}(\frac{f}{f_{base}} - 1)\right) \leq 1$ such that this quantity can be interpreted as a buy-up probability. In fact, the term can be reinterpreted as a willingness to pay distribution from which each customers individual willingness to pay is drawn. This choice of reference price is therefore used for generating customers in the simulation.

Numerical stability considerations, on the other hand, suggest to use a reference price that is in the middle of the overall price range. This is therefore done during estimation. To compare results of the estimation methods to the actual demand parameters, the parameters have to be converted to the same reference price f_{base}. When changing the reference price f_{base} to f'_{base}, the new demand parameters are

$$x'_{volume} = x_{volume} \cdot \exp\left(-x_{elast}(\frac{f'_{base}}{f_{base}} - 1)\right) \tag{5.5}$$

$$x'_{elast} = x_{elast} \cdot \frac{f'_{base}}{f_{base}} \tag{5.6}$$

The current demand model of REMATE does not allow for the exact customer behavior described here. Therefore, we have extended REMATE such that it can generate Hybrid demand, that changes over the booking horizon according to equation 5.1 and between departures according to the time evolution equation 4.9.

5.4. Approximating Demand Parameters from Real World Data

Aside from defining the demand model itself, its parameters need to be set as well, such that realistic demand can be generated in the simulation. Therefore, the goal of this section is to find realistic demand parameters for use in this simulation study. However, we're not aiming at estimating the demand parameters from real world booking and availability data. First, this would require using one of our proposed estimation methods, the quality of which we cannot ascertain before conducting this simulation study. So using these methods may lead to a circular argument. Second, acquiring the data on the necessary level of detail and with high quality is a difficult undertaking in itself. Last, and most importantly, good estimation methods should work for any set of demand parameters in a realistic range. Therefore, we focus on finding this realistic range from readily available airline data, with special attention to how fast these parameters change over time.

5.4.1. Analysis of Yield Data

As part of the research cooperation with Lufthansa German Airlines, we have access to twelve-year time-series (2001-2012) of monthly yields (average revenue per passenger) and monthly passenger counts per origin-destination pair for the ten largest [2] domestic, continental and intercontinental routes. Yields are measured from the passengers' viewpoint, since the goal is to explain passengers' behavior with the help of this variable. Yields from the airline perspective are usually lower, since fees and taxes received from customers are immediately passed on to other entities, and as such are irrelevant to the airline's internal accounting.

We first discuss how to find an approximate value for price elasticity x_{elast} and its lag-1 variance $var(x_{elast,t} - x_{elast,t-1})$. We have to make the assumption that all customers are price-sensitive, that the yield is a proxy for the lowest available price and that this lowest available price was optimal. It can easily be derived that the optimal price for price-sensitive demand only is $p^* = \frac{f_{base}}{x_{elast}}$. Consequently, we find that $x_{elast,t} = \frac{f_{base}}{\text{Yield}_t}$ where f_{base} is the reference price of our choice. The mean of the $x_{elast,t}$, provides us with a rough estimate of the true elasticity parameter. Furthermore, we can analyze how fast $x_{elast,t}$ changes over time. Seasonal effects are usually present in airline data and real-world forecasters usually have dedicated modules to handle seasonalities. This is not in the scope of this work however, so seasonal effects have to be removed from the data beforehand. We use the stl() function of the R programming language (R Core Team 2012) to accomplish this. This yields a new times series y_t which equals time series $x_{elast,t}$, but with seasonal effects removed.

We assume that we do not observe the price elasticity perfectly, but that there is

[2] as measured by total revenue in the complete time period

random observation noise. The state-space model is as follows:

$$x_t = x_{t-1} + w_t \qquad w_t \sim N(0, \sigma_w^2) \tag{5.7}$$

$$y_t = x_t + v_t \qquad v_t \sim N(0, \sigma_v^2) \tag{5.8}$$

Both, σ_v^2 and σ_w^2 are unknown, but for our purposes we are only interested in the latter. The variances of the lag-k differences $var(y_t - y_{t-k}) = 2\sigma_v^2 + k\sigma_w^2$ (called "lag-k variances" in the remainder) are a linear function of k in this model and the variance of interest σ_w^2 is the slope of this function. A simple linear regression over the lag-k sample variances therefore yields an estimate for σ_w^2. The deviation from linearity in the actual data will also be a hint at how closely the data match the state-space model above.

In the simulation, variances are specified as relative standard deviations and describe the change between consecutive departures and not between months. As such, the value used in the simulation and reported in this section is

$$\sigma_{rel} = \frac{\sqrt{\hat{\sigma}_w^2}}{4 \cdot \frac{1}{T} \sum_t y_t}. \tag{5.9}$$

Figure 5.1 shows the lag-k sample variances for the ten largest intercontinental routes in the Lufthansa network. Almost all curves are nearly linear for $k \leq 10$, after that the slope of some of the curve decreases, and in many cases the variances will start to fall again. This suggests that in the short term, within about one year, price elasticities evolve in agreement with our model. In the longer term however, there is a negative autocorrelation that reduces the lag-k variances for large k. Although the exact reason for this remains unclear, one possible explanation is that airlines will adjust their schedule and capacities to long-term demand changes. This would create a negative feedback loop on average prices and thus explain the long-term negative autocorrelation observed here. We therefore conjecture that these effects are not driven by demand changes but by adjustments of the supply, which can be ignored for our purposes. The respective charts for continental and domestic routes show qualitatively the same picture and have been referred to the appendix, section 10.2.

The almost linear parts of the lag-k variance curves yield an estimate of the variance σ_w^2, as described above. A criterion based on the r^2-measure of the linear regression determines the cut-off k below which each curve is regarded as linear. Figure 5.2 shows a box plot and a point cloud of the estimates elasticity parameters x_{elast} and their respective relative standard deviations.

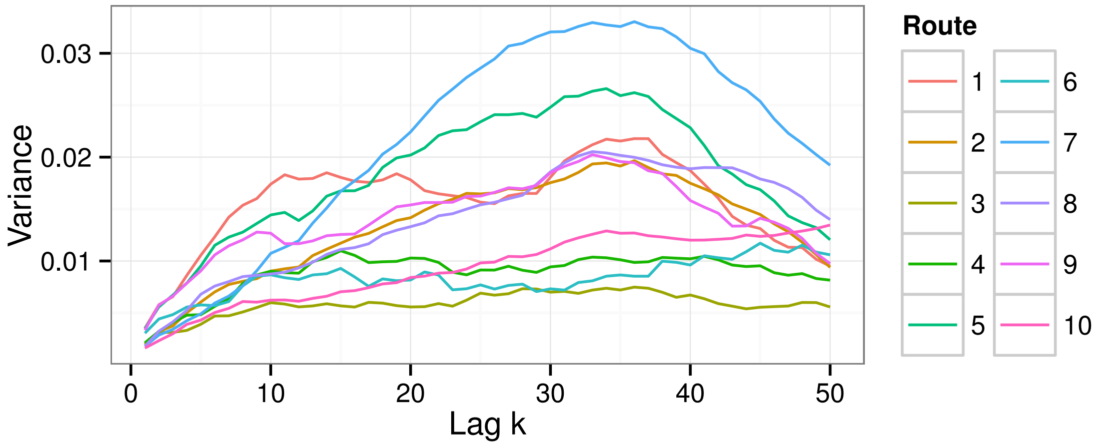

Figure 5.1.: Lag-k variances of elasticity estimate for the ten largest intercontinental routes; $f_{base} = 500$

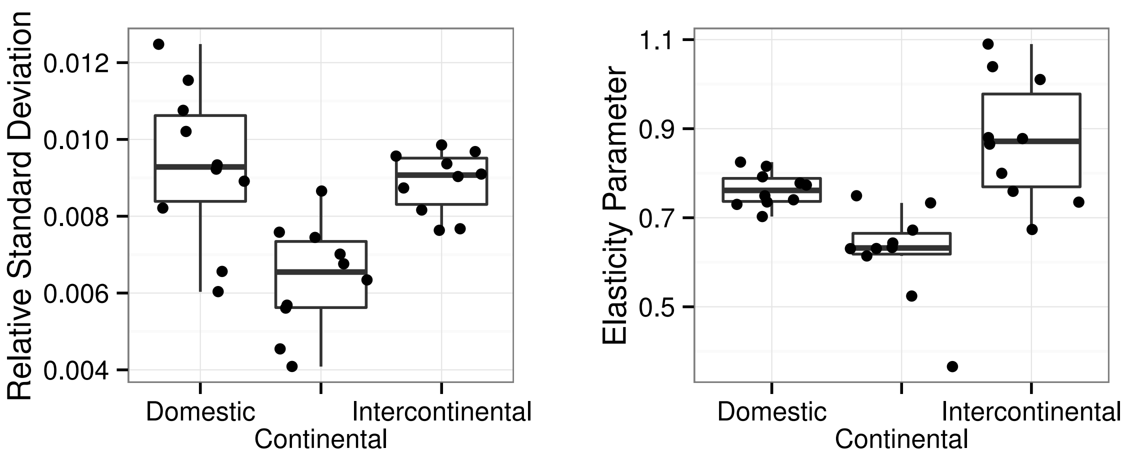

Figure 5.2.: Boxplot and point cloud of the estimated elasticity parameter x_{elast} and its relative standard deviation; $f_{base} = 500$ for intercontinental routes and $f_{base} = 100$ for all others

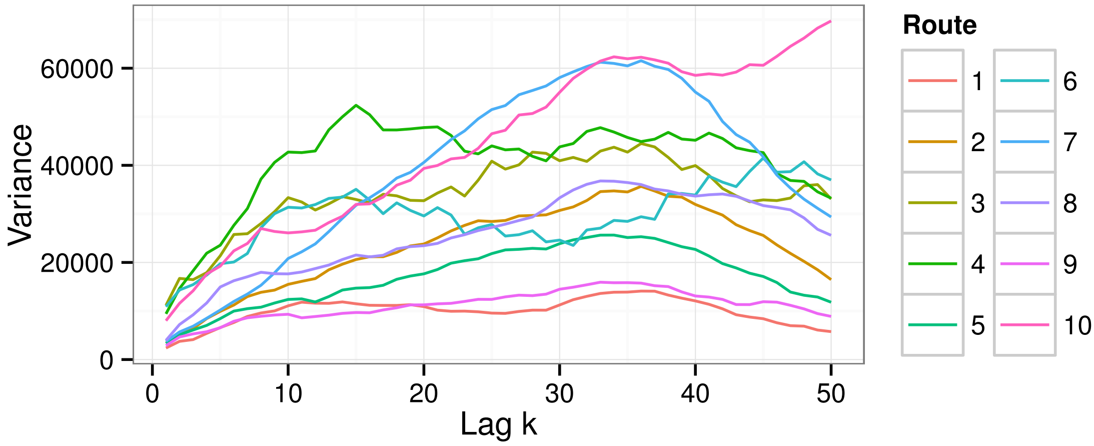

Figure 5.3.: Lag-k variances of volume estimate for the ten largest intercontinental routes; $f_{base} = 500$

5.4.2. Analysis of Passenger Count Data

The analysis of Passenger Count Data focuses on the relative standard deviation of the passenger counts. The overall volume itself has to be adjusted to the capacities used in the simulation and is calibrated to meet realistic seat load factors (= number of passengers / number of seats) instead.

The observed volumes are a function of the base volume x_{volume}, which is the focus of this analysis, and the buy-up probability given by the customer choice function. Under the same assumptions as in the previous section, the optimal price p^* can be substituted into the customer choice function, resulting in the buy-up probability $\exp(x_{elast} - 1)$. We therefore divide the observed passenger counts by this buy-up probability, using the estimated $\hat{x}_{elast}$ for each route and month from the previous step. The resulting time series are processed as in the previous section. Figure 5.3 shows the lag-k sample variances of the volume estimate x_{volume} for the ten largest intercontinental routes. Again, for $k \leq 10$, we observe almost linear behavior for nearly all routes, and for higher k the slopes of more and more curves start to level off. The interpretation is analogous to the one for the elasticity estimates. Figure 5.4 is a box plot of the estimated relative standard deviation of the volume parameter x_{volume}.

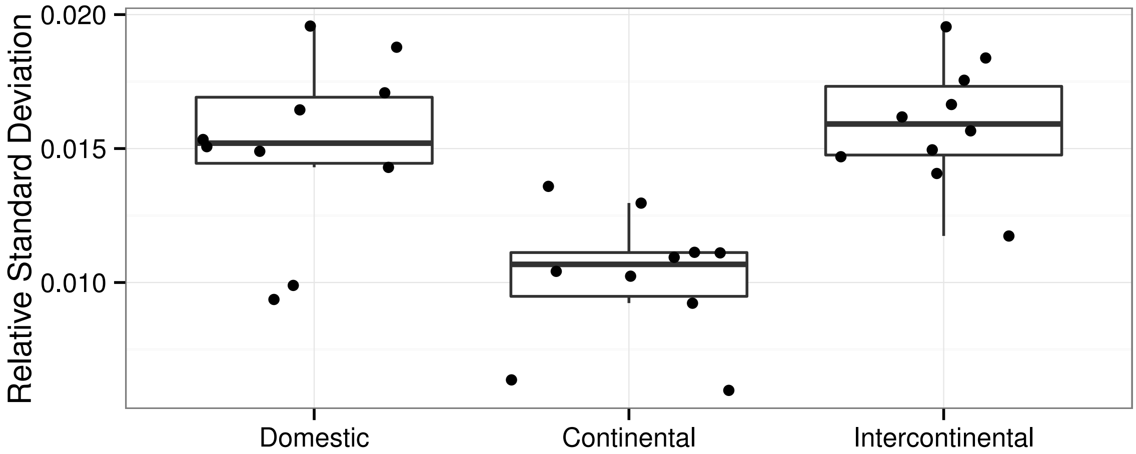

Figure 5.4.: Boxplot and point cloud of the estimated relative standard deviation of the volume parameter x_{volume}; $f_{base} = 500$ for intercontinental routes and $f_{base} = 100$ for all others

5.4.3. Price Elasticity over the Booking Horizon

The customer choice model in this simulation uses a particular function to model the change of the price elasticity parameter over the booking horizon (see section 5.3, equation 5.1). Under the assumptions from above, the trajectory of real-world yields over the booking horizon can be used to estimate the price elasticity parameter over the booking horizon. Figure 5.5 shows a comparison of estimated price elasticities and fitted values, according to equation 5.1. From a visual inspection, the fit seems quite good for all three curves. We were unable to achieve comparable results with fewer degrees of freedom or other functional forms.

5.4.4. Number of Price-Sensitive Customers

The relative shares of price-sensitive customers and product customers per booking class is another required input for the simulation. The assumption is that bookings in published fares are price-sensitive, while bookings in corporate, bulk, or target group fares are product customers. These customers don't see the regular, published fares, so it would not be adequate to use the price-sensitive buy-up model which is based on published fares. Furthermore, for these customer groups fares usually exist only in a subset of booking classes, such that full buy-down or buy-up behavior is not realistic.

Thus, we use fare type data to define the number of product customers in relation to price-sensitive customer, keeping in mind that both parts of the assumption are certainly

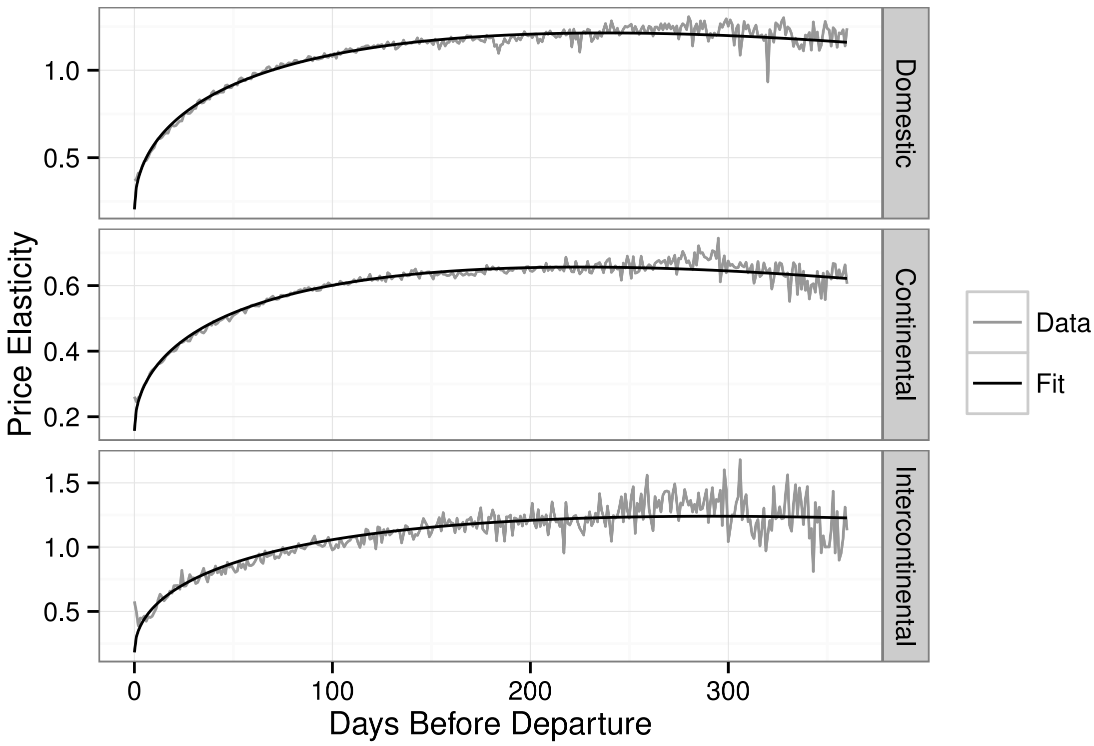

Figure 5.5.: Plot of estimated price elasticity parameter and fitted values over the booking horizon; $f_{base} = 500$ for intercontinental routes and $f_{base} = 100$ for all others; ten largest routes by revenue per region

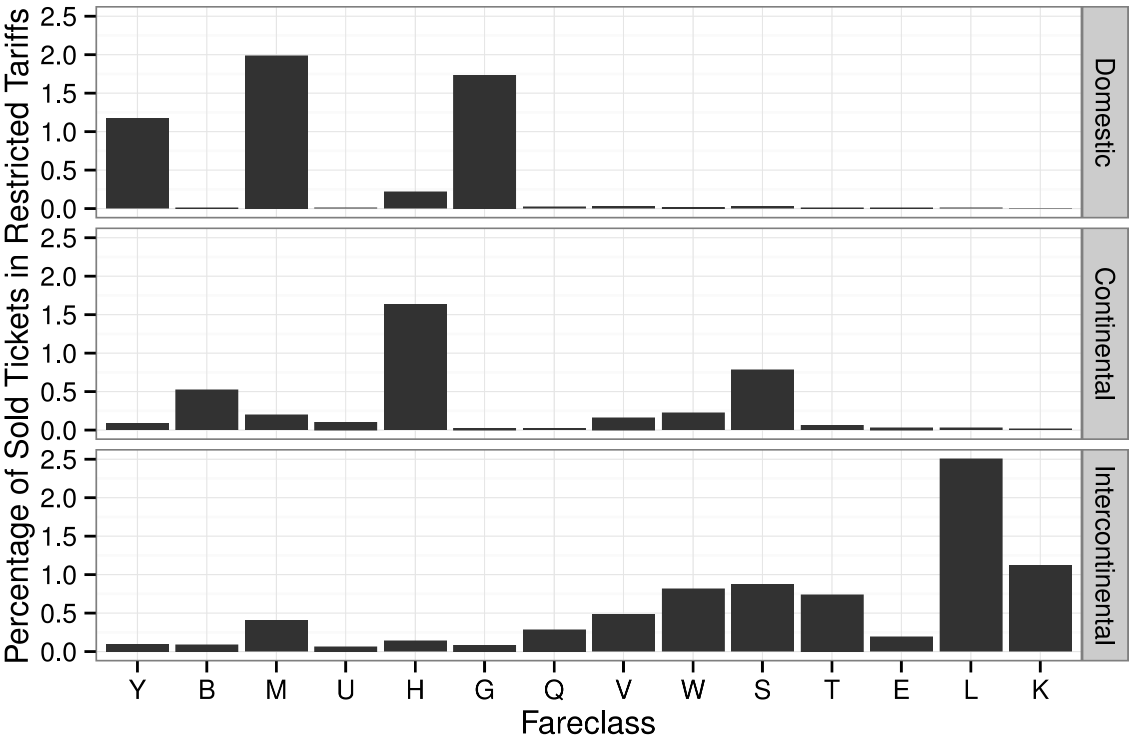

Figure 5.6.: Sold tickets in corporate, bulk and target group fares per fareclass as a percentage of all sold tickets on all routes in the dataset of the respective region

not entirely true: not all customers of published fares are completely price-sensitive and not all customers in special groups are only interested in a single booking class.

Figure 5.6 shows the number of sold tickets with special fares as a percentage of all sold tickets in each region. In the simulation, these percentages are used to set the number of product customers as a fraction of overall demand.

5.5. Summary

As mentioned at the beginning of this section, the results of this data analysis should be regarded as very rough estimates of the real demand parameters due to the set of restrictive assumptions that were necessary. Nevertheless, the shape of the lag-k variances offers some justification for the model used and the values of the final estimates seem robust in that they are all within the same order of magnitude. Moreover, the general level of price-elasticities and their path over the booking horizon roughly agrees

Property	**Domestic**	**Continental**	**Intercontinental**
Volume: Rate of Change	High	Low	High
Elasticity:			
Value	High	Medium	Low
Rate of Change	High	Low	High
Change over Booking Horizon	High	Low	High
Product Demand: Volume	Low	Low	High

Table 5.1.: Overview of the qualitative properties of the three simulation scenarios Domestic, Continental and Intercontinental.

with the values reported by Gwartney et al. (2008, p. 429), which are comparable to our values at the base price. We are therefore confident that the results fulfill the initial goal of finding parameters in a realistic range.

Table 5.1 summarizes the qualitative differences between the three scenarios Domestic, Continental and Intercontinental. Note that elasticity values are only directly comparable if they refer to the same base price. For the ratings given here, all elasticity values were converted to the same base price and are therefore not in the same order as reported in figure 5.2. The overall demand volume is a separate parameter and is therefore not reported in this table. Many more combinations of parameters are obviously possible, however, for our simulation study we had to restrict ourselves to this small set.

6. New Methods for Demand Estimation

Once an airline has settled for a particular demand model H_a (see chapter 4), the question arises of how to set the parameters x of that demand model. The airline only observes sales numbers, and the relationship between those and the underlying demand parameters is often indirect – at least for choice-based demand models – and noisy. Moreover, demand is also expected to change over time, such that parameter estimation has to be a continuous effort, embedded into the overall revenue management process.

This chapter proposes two novel approaches to demand estimation, based on the state-space representation developed in chapter 4. Additionally, we introduce a heuristic for active estimation that deliberately introduces price fluctuations to improve the estimation of price elasticities. Using the simulation model from chapter 5, we then compare our proposed methods to a set of benchmark methods.

6.1. Unscented Kalman Filter

6.1.1. Background

For state-space models with linear state transition and observation functions and with only additive Gaussian noise, the estimation problem can be solved analytically to optimality. If in the state-space model of equations 4.12 and 4.13, the demand function H_a is required to be linear, we have just such a simple model. The analytical solution to the demand estimation problem is called the Kalman Filter (Kalman 1960), and has a number of desirable properties: The Kalman Filter

- is the minimum mean-squared error (MMSE) estimator and it attains the PCRB exactly.
- is computationally fast

- can be computed recursively, that is only the last estimate $\hat{x}_{t-1}$, its covariance P_{t-1} and the current observation b_t is required to produce the next estimate $\hat{x}_t$ and its covariance P_t. Observation data can be discarded after it was used once in the estimate update.

- produces not only a point estimate, but the complete and exact posterior distribution of x_t – given the observation history – by providing the covariance of the estimate P_t. This may be useful in optimization, e.g. to explore the state-space in the direction of greatest uncertainty.

Due to these advantages, the Kalman Filter is used in a wide array of applications, including dynamic pricing (Kwon et al. 2009). However, it is by itself not applicable to the original model from equations 4.12 and 4.13, since in general and in our particular simulation setting, the demand function H_a is *not* linear. This is in contrast to the model of Kwon et al. (2009) who employ a linear demand function and thus were able to use the Kalman Filter directly. For our purposes, the Kalman Filter has to be extended to cover the case of a non-linear observation function.

Starting with the original paper of Kalman (1960) a tremendous amount of work has been put into extending the model in various directions and adapting it to numerous areas of application. Here, we just give an overview of the articles that are of particular relevance to our work.

The so-called Extended Kalman Filter is the straight-forward extension of the Kalman Filter to non-linear models, using a first-order, linear approximation of the original model (Wan & Merwe 2000). It addresses non-linearity, both in the state evolution function as well as in the observation function. Julier & Uhlmann (1997) propose a different approximation based on an alternative parameterization of the normal distribution, this approach is called the Unscented Kalman Filter (UKF). While still conceptually simple and computationally efficient, this method outperformed other methods for overcoming the linearity restriction, specifically the Extended Kalman Filter, in the numerical experiments of Julier & Uhlmann. For this reason, we decided to use the Unscented Kalman Filter for our setting. Julier later extends his results, introducing a scaling factor and correction term that accounts for non-linearities above the second order (Julier 2002). Our version of the Unscented Kalman Filter is based on the formulation of Wan & Merwe (2000) who further extend the results of Julier to a more general class of filtering problem.

As noted in chapter 4, equations 4.12 and 4.13 also introduce the assumption of Gaussian error terms, which in turn result in potentially negative and non-integer booking predictions. There are extensions to the Kalman Filter that address these shortcomings.

The issue of non-negativity is treated by Simon (2010) who considers Kalman filtering under state constraints. These constraints can be set up such that the *expected* observations cannot be negative. However, due to the Gaussian measurement error, individual observation could still be negative. Brankart (2006) extends the Kalman Filter with linear inequality constraints and shows that an optimal, MMSE estimator can still be derived if the error terms are assumed to be drawn from truncated Gaussian distributions.

Finally, the case of integer-valued measurements or quantized measurement is considered by Duan et al. (2008) and Karlsson & Gustafsson (2005). The introduction of quantized measurement complicates the estimation tremendously, Duan et al. employ a sophisticated approximate numerical approach while Karlsson & Gustafsson resort to a Particle Filter, which is essentially a Monte Carlo method for state estimation. The quantization model in both cases corresponds to rounding the outcomes of the observation function before they can be observed. This seems to be a somewhat unnatural assumption compared to, say, assuming a Poisson distribution for the booking data.

Due to the limitations of these approaches and their significant additional complexity, we did not include any of them in our proposed Unscented Kalman Filter for demand estimation. Instead, we will directly incorporate the assumption of Poisson bookings in our second proposed demand estimation method in section 6.2.

6.1.2. Unscented Kalman Filter for Demand Estimation

The Unscented Kalman Filter proposed by Julier & Uhlmann (1997) and extended by Wan & Merwe (2000) can handle non-linearities both in the state evolution equation and the observation equation. Here, we only make use of the latter since the state evolution equation (eqn. 4.10) of the original model is already linear. The UKF is a heuristic, i.e. an inexact method. Therefore, it is not necessarily the MMSE estimator.

We formulate the UKF for a generalized Hybrid choice function that can be decomposed into a linear part L_a and a non-linear part H_a^N:

$$H_a(x) = L_a x^L + H_a^N(x^N) \tag{6.1}$$

where $x = (x^L, x^N)$. Separating out the linear part of H_a allows for much more efficient implementation by essentially combining a regular Kalman Filter for the linear part with the UKF for the non-linear part. In case the demand function is not separable, we can simply set $L_a = 0$. In the other extreme, if the demand function is linear such that $H_a^N(x^N) = 0$, our formulation is identical to the original Kalman Filter.

The UKF assumes Gaussian, though not necessarily additive, observation errors. Thus, we implicitly assume the following observation equation:

$$b_t = L_a x_t^L + H_a^N(x_t^N) + v_t \qquad\qquad v_t \sim N(0, diag(H_a(x_t))) \tag{6.2}$$

where $diag(H_a(x))$ is the matrix with $H_a(x_t)$ on its diagonal and all other entries zero. Thus, the variance of the observation equals its expectation, consistent with a Poisson distribution.

At every time step t, the UKF will produce an estimate $\hat{x}_t$ for the demand parameters and a covariance matrix P_t that together define the (approximate) current belief about the real demand parameters. In other words, $\hat{x}_t$ and P_t are the parameters of the (approximate) posterior distribution of x_t, given all observations up to time t.

Our formulation of the UKF is a direct specialization of that given by Wan & Merwe (2000). We can simplify the computation of the update equations somewhat by exploiting the linearity of the state-space evolution equation and the decomposition of H_a into a linear and non-linear part. As such, our version gives identical numerical results to what the algorithm of Wan & Merwe (2000) would produce, however with reduced computational effort. In our simulation setting, this reduction in computational effort was necessary to keep running times practicable. Here, we only reproduce our final formulation of the UKF. The derivation from the equations of Wan & Merwe (2000) is given in the appendix, section 10.1.2.

The algorithm is as follows: First, decompose[1] the covariance matrix P_t into an upper triangular matrix U such that $UU^T = P_t$. From this, compute the set of $2n^N + 1$ sigma points σ_i, where n^N is the number of demand parameters of the non-linear part of the choice function:

$$\sigma_0 = \hat{x}_t \tag{6.3}$$

$$\sigma_i = \hat{x}_t + \sqrt{n+\kappa}\, U_i \qquad\qquad i = 1, \ldots, n^N \tag{6.4}$$

$$\sigma_i = \hat{x}_t - \sqrt{n+\kappa}\, U_{i-n^N} \qquad\qquad i = n^N + 1, \ldots, 2n^N \tag{6.5}$$

[1]See section 10.1.1 in the appendix on how to compute this decomposition.

Here, U_i is the i-th column of U, n is the number of demand parameters for the complete demand function and $\kappa = \alpha^2 \cdot n - n$. Parameter α is a scaling parameter that determines the distance of the sigma points from their mean $\hat{x}$. Additionally, in equation 6.10, there is a second parameter, β, which can be used to incorporate knowledge of the distribution of $\hat{x}$. We use the values $\alpha = 10^{-3}$ and $\beta = 2$ as suggested by Wan & Merwe (2000).

Next, apply the non-linear part of the choice function to each sigma point to obtain

$$g_i = H_a^N(\sigma_i) \qquad\qquad i = 0, \ldots, 2n^N. \tag{6.6}$$

From this, compute the expected bookings from the non-linear part z^N, the linear part z^L and their sum z:

$$z^N = \frac{\kappa + n^L}{\kappa + n} g_0 + \frac{1}{2(n+\kappa)} \cdot \sum_{i=1}^{2n^N} g_i \tag{6.7}$$

$$z^L = L_a \hat{x}_t^L \tag{6.8}$$

$$z = z^L + z^N \tag{6.9}$$

Additionally, find the booking covariance matrix from the non-linear part

$$\begin{aligned} P_{z^N z^N} = & \left(\frac{\kappa + n^L}{\kappa + n} + (1 - \alpha^2 + \beta) \right) \cdot (g_0 - z^N)(g_0 - z^N)^T \\ & + \frac{1}{2(n+\kappa)} \cdot \sum_{i=1}^{2n^N} (g_i - z^N)(g_i - z^N)^T \end{aligned} \tag{6.10}$$

and the cross-covariance between the complete state x_t and the non-linear bookings z^N

$$P_{xz^N} = \frac{1}{2(n+\kappa)} \cdot \sum_{i=1}^{2n^N} (\sigma_i - \hat{x}_t)(g_i - z^N)^T. \tag{6.11}$$

Let P^L be the upper left $n^L \times n^L$ block of P_t and $P_{x^L z^N}$ be the first n^L rows of P_{xz^N}. Then, compute the total booking covariance

$$P_{zz} = L_a P^L L_a^T + P_{z^N z^N} + 2 \cdot L_a \cdot P_{x^L z^N} + diag(H_a(\hat{x}_t)) \tag{6.12}$$

and with the left $n \times n^L$ block P^{NL} of P_t, construct the total cross-covariance

$$P_{xz} = P^{NL} L_a^T + P_{xz^N}. \tag{6.13}$$

The values obtained for z, P_{zz} and P_{xz} are now used for the regular Kalman Filter update and predict equations. Hence, we compute the Kalman gain $K = P_{xz}P_{zz}^{-1}$ and from that the new demand estimate

$$\hat{x}_{t+1} = \hat{x}_t + K \cdot (b_t - z) \tag{6.14}$$

$$P_{t+1} = P_t - KP_{zz}K^T + Q \tag{6.15}$$

Improved performance in this formulation results from fewer evaluation of the demand function. In the original version, there would be $2n + 1$ sigma points, each resulting in one evaluation of the demand function, whereas in our specialized version there are only $2n^N + 1$ sigma points. In one of our simulation settings, there are $n = 333$ parameters, with $n^N = 25$ parameters of the non-linear part, such that our formulation reduces the number of demand function evaluations from 667 to 51. Depending on the complexity of the demand function, this can have a large impact on overall performance. The remaining parts of the algorithm remain unchanged in complexity, since they are dominated by the $\mathcal{O}(n^3)$ Cholesky decomposition.

6.2. Particle Filter

6.2.1. Background

The Particle Filter (PF) is a Monte-Carlo approach to state-space estimation. It has the advantage that no assumptions have to be made on the dependence of x_t on x_{t-1}, nor on the form of the observation function, nor on the distribution of the noise terms. In this general setting the belief distribution of state x_t has no closed-form. The Particle Filter solves this by approximating the continuous belief distribution by a large number of randomly chosen, discrete points. Each of these so-called *particles* has a location in state space and a weight that represents the likelihood of this particle explaining the observations.

Handschin & Mayne (1969) propose this technique for the first time, however not using the term Particle Filter yet. Doucet et al. (2000) review numerous Particle Filter methods proposed in the literature and develop a general framework. They note that the various Particle Filter variants from the literature primarily differ in the choice of importance weighting function. This importance function guides the creation of particles, such that most particles are in areas of high interest, i.e. where the belief distribution has non-negligible probability density. Moreover, Doucet et al. (2000) argue that "one obtains poor performance when the importance function is not well chosen".

We therefore investigate two variants of importance function in this thesis. The first uses a fixed and easy to evaluate importance function. It was used in the original work of Handschin & Mayne (1969), and – among others – by Gordon et al. (1993) and Kitagawa (1996). Our method is closely related to that of Gordon et al. (1993), since not only the same importance function is used, but also the same re-sampling algorithm. Re-sampling becomes necessary when the particle set degrades such that most weight is concentrated on very few particles. While a well chosen importance function will slow down this process, it will eventually happen regardless of the importance function, as shown by Doucet et al. (2000) using a result of Kong et al. (1994). The re-sampling method Gordon et al. (1993) and by extension we use is based on a result by Rubin (1988).

The second variant we evaluate aims to approximate the optimal importance function, which was introduced by Zaritskii et al. (1975). This function has no analytical form for our model, so we use a local linearization as proposed by Doucet et al. (2000) to sample from it. We use the same resampling method as above, but resampling is less frequent here, since the optimal importance function minimizes the particle weights' variance (Doucet et al. 2000).

The Particle Filter is a more general extension of the Kalman Filter, in that it can handle almost any state-space model. This flexibility however comes at the price of computational effort. The Particle Filter starts with a large number of hypotheses (=particles) about the real parameters and updates the likelihood of these hypotheses as new observations arrive. The treatment of a large number of particles makes this approach computationally more demanding. Conceptually however, the method is very straight-forward and easy to implement.

The Particle Filter method is closely related to Monte Carlo integration methods with importance sampling. Therefore, this type of filter is also known as a Sequential Importance Resampling (SIR) Filter. Asymptotically, when the number of particles tends to infinity, the Particle Filter is the minimum mean squared error filter and the approximated posterior density converges to the real posterior density (see e.g. Gordon et al. 1993). However, for a finite number of particles, little can be said about the quality of the filter and thus the required number of particles can only be found in experiments. In practice, there will be a trade-off between filter quality and required computational resources. It is noteworthy though, that the Particle Filter is very well suited for highly parallel implementation.

6.2.2. Particle Filter for Demand Estimation

Since the Particle Filter puts no restrictions on the state-space model, we use our original formulation of the state-space model, repeated here for the reader's convenience:

$$x_{t+1} = x_t + w_t \qquad\qquad w_t \sim N(0, Q) \tag{6.16}$$

$$b_t \sim Poi(H_a(x_t)) \tag{6.17}$$

Let N be the number of particles used. At every time step, the Particle Filter holds a set of particles $\mathcal{P}_t = \{\hat{x}_{tk}, k = 1, \ldots, N\}$ and a corresponding set of weights $\mathcal{W}_t = \{\omega_{tk}, k = 1, \ldots, N\}$. Each particle represents a potential parameter vector for the demand function H, and its corresponding weight is the likelihood of that parameter vector being the true parameter vector.

The weights and particles together form a discrete distribution that approximates the actual continuous posterior distribution. The expected value for the parameter estimate at time t can be computed as the mean of the particles at time t, such that $\hat{x}_t = \frac{1}{N} \sum \omega_{tk} \hat{x}_{tk}$.

When a new set of observed bookings b_t and availabilites a_t arrives, a new set of particles is generated and their respective likelihoods are evaluated. The new set of particles could theoretically be drawn from a uniform distribution over the whole parameter space. This, however, would lead to large number of particles with very small likelihoods, and therefore unnecessarily large computation time and memory requirements. The key idea in importance sampling is to put most particles in regions of "high interest". These regions are described by a so-called importance function $\pi(x|x_{0:t-1,k}, b_{0:t})$, which assigns a weight to each point in the parameter space, based on the particle's past trajectory and the history of observations. There are multiple choices for the importance function, and most variants of the Particle Filter found in the literature differ primarily in this choice (Doucet et al. 2000).

Regardless of the choice of importance function, the general algorithm is as follows:

- For $i = 1, \ldots, N$: Sample $\hat{x}_{tk} \sim \pi(x|\hat{x}_{0:t-1,k}, b_{0:t})$
- For $i = 1, \ldots, N$: Compute importance weights

$$\omega'_{t,k} = \omega_{t-1,k} \cdot \frac{p_a(b_t|\hat{x}_{tk}) \cdot p(\hat{x}_{tk}|\hat{x}_{t-1,k})}{\pi(\hat{x}_{tk}|\hat{x}_{0:t-1,k}, b_{0:t})} \tag{6.18}$$

- For $i = 1, \ldots, N$: Normalize importance weights $\omega_{t,k} = \frac{\omega'_{t,k}}{\sum \omega'_{t,k}}$

The conditional probability $p_a(b_t|\hat{x}_{tk})$ is given by the Poisson probability distribution function and the choice function H:

$$p_{a_t}(b_t|\hat{x}_{tk}) = \prod_i \frac{(h_{a,i}(\hat{x}_{tk}))^{b_{t,i}}}{b_{t,i}!} \cdot e^{(h_{a,i}(\hat{x}_{tk})} \tag{6.19}$$

The conditional probability $p(\hat{x}_{tk}|\hat{x}_{t-1,k})$ describes the state evolution. From equation 6.16, we find that this is a multivariate Gaussian distribution with mean $\hat{x}_{t-1,k}$ and covariance Q.

We evaluate two different choices for the importance function. The first one is the one-step update distribution $p(x_t|x_{t-1})$ which was used by Gordon et al. (1993). The advantage of this choice is that in our model this is just a multivariate Gaussian distribution from which samples can be drawn easily. Additionally, the update equation for the particle weights simplifies to $\omega'_{t,k} = \omega_{t-1,k} \cdot p_a(b_t|\hat{x}_{tk})$, making it one of the computationally most efficient variants. Doucet et al. (2000) note however that this choice of importance function "is often inefficient in simulations as the state space is explored without any knowledge of the observations."

As mentioned above, it is desirable to keep the particle weights as evenly distributed as possible. This can be measured by the particle weights' variance. It can be shown that the importance function $\pi(x|x_{0:t-1,k}, b_{0:t}) = p_a(x|x_{t-1,k}, b_t))$ minimizes the variance of the particle weights (Doucet et al. 2000). It is therefore called the optimal importance function and is the base of the second candidate importance function. In our model however, sampling from this function is not possible analytically. As proposed by Doucet et al., we therefore approximate the importance function locally around x by a multivariate Gaussian distribution. Towards this, define the log-likelihood $l(x) = \log p_a(x|x_{t-1}, b_t)$. We can then compute the first two derivatives

$$\begin{aligned} l(x) &= \text{const.} - \frac{1}{2}(x - x_{t-1})^T Q^{-1}(x - x_{t-1}) \\ &\quad + \sum_{i:h_i(x)>0} (b_{ti} \log h_i(x) - h_i(x)) \end{aligned} \tag{6.20}$$

$$\nabla_x l(x) = -Q^{-1}(x - x_{t-1}) + \sum_{i:h_i(x)>0} \nabla_x h_i(x) \cdot (\frac{b_{t,i}}{h_i(x)} - 1) \tag{6.21}$$

$$\frac{\partial^2}{\partial x \partial x^T} l(x) = -Q^{-1} + \sum_{i:h_i(x)>0} \left(\frac{\partial^2}{\partial x \partial x^T} h_i(x) \cdot (\frac{b_{t,i}}{h_i(x)} - 1) - \Delta_x^x h_i(x) \cdot \frac{b_{t,i}}{h_i^2(x)} \right) \tag{6.22}$$

A second order Taylor expansion yields the covariance $\Sigma = -l''(x)^{-1}$ and mean $m =$

$x + \Sigma \cdot l'(x)$. The point x around which the log-likelihood function is approximated locally should be the mode of $p_a(x|x_{t-1}, b_t)$ which can be found numerically with Netwon's iterative method. Constructing this importance function is computationally expensive when the number of parameters becomes large. In our setting it is prohibitively expensive to compute individually for each particle, which is implied by $p_a(x|x_{t-1,k}, b_t))$. Instead we compute the covariance matrix and mean displacement vector once for the average $\hat{x}_{t-1}$ of the current set of particles, and then sample from that distribution for each particle x_{tk}. Still, this variant is computationally more expensive than the first, yet it does have the advantage of putting more weight in areas of the parameter space that are in agreement with the current observation.

In principle, this already describes a working Particle Filter. In real-world implementations, however, the particle cloud will degenerate at some point, as analytically proven by Doucet et al. (2000). That is, most of the particle weight will be concentrated on a single particle. To overcome this, the particles have to be resampled from time to time. We employ the resampling strategy presented by Doucet et al.. If the estimated number of effective particles $N_{eff} = \frac{1}{\sum \omega_{tk}^2}$ is smaller than some minimum fraction of the total number of particles N, the particle weights are degenerated and resampling should be performed.

During resampling, each particle $\hat{x}'_{t,k}$ is replaced by an existing particle $\hat{x}_{tr}$, where the index r is drawn randomly with replacement from the set $1, \ldots, N$ with probabilities ω_{tk}. The new weights are $\omega'_{t,k} = \frac{1}{N}$. It can easily be verified that the moments of the particle distribution remain unchanged in expectation by this operation. Furthermore, the number of effective particles N_{eff} now equals the total number of particles N.

The simulation results in section 6.5.1 compare the two Particle Filter variants described here, as well as two different sizes N of the particle cloud.

6.3. Active Estimation

Up to this section, we have ignored the revenue management feedback loop described in section 4.1.4. This is not necessarily a valid assumption. The revenue impact of methods for active learning, which aim to include the effect of an improved forecast on future revenues, varies between minor (Aviv & Pazgal 2005) and very high (Carvalho & Puterman 2004) in the literature. We therefore evaluate our own heuristic for active learning that we call Active Estimation.

6.3.1. Background

The dynamics of information acquisition and loss over time is the basis of active learning approaches. Apart from Aviv & Pazgal (2005) and Carvalho & Puterman (2004) from the dynamic pricing literature, there is a number of articles form inventory management that also consider informational dynamics and active learning. Lariviere & Porteus (1999) and Ding et al. (2002) consider the newsvendor problem and show that the optimal stocking level with active learning is higher than the myopic stocking level. Since demand above the stocking level is censored and thus unobserved the increased stocking level provides additional information about the demand distribution which leads to better inventory decisions in subsequent periods. Chu et al. (2008) show how to compute the optimal stocking level for a wide range of demand distributions.

Kiefer & Nyarko (1989) analyze a firm that learns the coefficients of a linear model from observed data and which controls the value of the regressor variable. Again, this value influences both current revenue and future rewards through improved information on the coefficients. Kiefer & Nyarko show under few additional assumptions that the optimal strategy typically differs from the myopic strategy that ignores the informational dynamics. Only in the limit of infinitely many observations, does the optimal strategy converge on the myopic strategy. The demand model that we wish to estimate in this thesis is not linear, however, and our model parameters evolve over time, such that it is unlikely that these structural results apply in our case.

In control theory, there are two streams of literature related to the active learning problem. One is that of measurement scheduling which aims at finding a set of costly measurements (from a set of candidates) that maximizes the Fisher information or some other measure of estimation quality within a budget constraint. See Shakeri et al. (1995) or Gupta et al. (2006) for example applications. However, measurement scheduling does not explicitly consider the trade-off between optimizing an objective function now and acquiring more information to improve that optimization in future periods. This trade-off is considered in the *dual control* problem. Wittenmark (1995) gives a problem definition and reviews various methods for solving it approximately.

In this thesis, we evaluate a heuristic that adds additional price variation specifically where demand estimates are highly uncertain. It is highly related to the simplest heuristic mentioned by Wittenmark (1995): *Perturbation signals.*

6.3.2. New Active Estimation Heuristic

With Active Estimation (AE), we exploit the information about estimate uncertainty provided by the filtering methods[2]. The goal is to selectively increase the variability of availabilities where demand uncertainty is high. Instead of using the best current demand estimate for optimization, we randomly draw a demand parameter vector from the uncertainty distribution. The intuition is, that when demand uncertainty is small, active learning is not required and we will draw demand parameters close to the current estimate. If, on the other hand, demand uncertainty is large, we draw demand parameters further away from the current estimate, which let the optimizer compute more strongly varying availabilities. The increased variability in availabilities improves demand estimation, reducing demand uncertainty in future runs.

While we always expect a reduction of mean-squared estimation error from using Active Estimation, the effect on revenue is less clear. In the short-term, Active Estimation reduces revenue by not using the current best estimate in optimization. However, this may be offset by an increase in revenue due to the increased forecast quality later on. The relative size of these effects determines the overall profitability of Active Estimation. We investigate this in a simulation study and report the results in sections 6.5.1 and 6.5.3.

6.4. Simulation Setup

The overall simulation model used here, has been described in chapter 5. For the simulation study in this chapter, we additionally implemented the Unscented Kalman Filter, the two variants of the Particle Filter and Active Estimation in REMATE.

6.4.1. Baseline Methods

To evaluate the performance of the proposed estimation algorithms, we need benchmark methods to compare against. The PCRB provides the best-case scenario in terms of mean-squared error, and optimization using real demand parameters yields the optimal expected revenue, the *true parameter revenue.* We extended REMATE, such that the PCRB and the true parameter revenue can be computed for any given simulation scenario. Additionally, we implemented three baseline demand estimation methods, that

[2]If an estimation method is used that does not yield uncertainty information, then the PCRB can serve as an approximation. Thus, Active Estimation is independent of any particular estimation and optimization method.

use a simple heuristic, a method from the literature, and maximum-likelihood estimation, respectively, to estimate demand parameters. These methods are explained in turn in this section.

Simple Estimation

In Simple Estimation (SE), each demand parameter is estimated independently. For each parameter, we find a value that is most consistent in least-squares sense with the current observations, holding all other parameters constant. Combining all these individual parameter estimates, yields a new parameter vector which obviously ”overshoots” the true parameters, since the change in each parameter alone could explain the observation. Exponential smoothing is used to avoid the overshoot. However, choosing a smoothing factor that works globally is difficult, here we used a value of 0.2, which is in the typical range of values used in revenue management applications (Talluri & Ryzin 2005, p. 437). Even with exponential smoothing, non-linear choice functions can produce extreme estimates. This is avoided by using a simple outlier detection that limits the maximum change of a parameter value between time steps.

Forecast Prediction

Forecast Prediction (FP) is an estimation method developed for the Passenger Origin-Destination Simulator (PODS) to forecast price-sensitive demand. Boyer (2010) and Guo (2008) describe both the simulator itself and the Forecast Prediction method. They report that this method performed best among the estimation methods implemented in PODS. We re-implemented Forecast Prediction in the REMATE simulation environment to compare it with our methods.

Forecast Prediction estimates the price-sensitivity parameter separately for each discrete time period of the booking horizon. In PODS, these individual estimates are subsequently smoothed using either linear regression or a logistic fitter on a transformed choice parameter (Boyer 2010). However, neither of these options is a good fit for our demand model in which the price-sensitivity parameter is a degree-two polynomial. Using an inadequate regression model might reverse any positive results from the Forecast Prediction method itself. A comparison to our methods might then be considered unfair, since our methods know the true functional form. Therefore, in our implementation of Forecast Prediction, we fit a degree-two polynomial to the price-sensitivity parameters estimated by Forecast Prediction using ordinary least-squares regression.

Forecast Prediction exclusively estimates price-sensitive demand. The product demand component is estimated in the same way as in Simple Estimation, using exponential smoothing with a learning rate of 0.2.

Maximum-Likelihood Estimation

A standard solution to this type of estimation problem is maximum-likelihood estimation (MLE). For example, Vulcano et al. (2012) use maximum-likelihood estimation, albeit resorting to an EM-algorithm to find the maximum, due to a more complex choice function. In our setting, it is feasible to numerically find the maximum of the likelihood function directly.

Given the joint probability $p_A(B,X)$, find a demand parameter trajectory $X = (x_0,\ldots,x_T)$ that maximizes $p_A(B,X)$ for the observed availability and booking histories A and B. From the original state-space model in equations 4.10 and 4.11, $p_A(B,X)$ can be decomposed into

$$p_A(B,X) = p(x_0)\prod_{t=1}^{T} p_{a_t}(b_t|x_t)\cdot p(x_t|x_{t-1}). \tag{6.23}$$

Maximizing equation 6.23 (or its logarithm) over X is a very high dimensional problem. As an example, in the scenarios of this simulation study, each individual x alone has dimension 333. After 100 simulation runs, the dimension of X will be $333\cdot 100 = 33300$. Maximizing any non-trivial function over that many parameters is a major challenge.

To make this problem more tractable, we limit the availability and booking histories to a rolling history of a fixed number of observations (here 25). Further, we assume that x remained constant within this limited observation history and that the initial x_0 is known and equals the estimate from the preceding time step, i.e. $x_0 = \hat{x}_{t-1}$. The joint probability function then becomes

$$p_A(B,x) = p(x|\hat{x}_{t-1})\prod_{t=1}^{T} p_{a_t}(b_t|x). \tag{6.24}$$

where the term $p(x|\hat{x}_{t-1})$ is given by the multivariate normal distribution with mean $\hat{x}_{t-1}$ and covariance matrix Q. The conditional probability $p_{a_t}(b_t|x)$ is the product of Poisson probability distribution functions with rates $\lambda_i = h_{a,i}(x)$. This yields the log-likelihood

$$\begin{aligned} L(x) = \log p_A(B,x) = &-\frac{1}{2}(x-\hat{x}_{t-1})^T Q^{-1}(x-\hat{x}_{t-1}) \\ &+\sum_{\tau}\sum_{i,h_i>0} b_{\tau i}\cdot\log\left(h_{i,a_\tau}(x)\right) - h_{i,a_\tau}(x) \\ &+ const. \end{aligned} \tag{6.25}$$

This function is maximized by finding a root of the first derivative with the iterative Newton method. $L(x)$ is not concave in general and therefore a global maximum is not guaranteed. In practice, this doesn't seem to be an issue, since the maximum is expected to be close to $\hat{x}_{t-1}$ which is therefore an excellent starting value.

Implementation Note Newton's method requires the first and second derivative of $L(x)$, $\nabla_x L(x)$ and $\Delta_x^x L(x)$ in the notation of section 4.4. A straight-forward implementation requires a large number of evaluations of the choice function H and its first two derivatives, a number that is linear in the length of the booking history. Evaluation of the Hessian matrices $\Delta_x^x h_i(x)$ is computationally expensive when the number of parameters is large and the booking vector is long. In the implementation we exploit the special structure of the hybrid choice function, namely that for fixed x

- product bookings in a class depend linearly and exclusively on the availability of that class, and
- price-sensitive bookings in a class depend linearly and exclusively on the amount of time for which that class was the lowest available.

Together, these two facts make it sufficient to evaluate the derivative and Hessian matrices of H once per iteration, independent of the length of the booking history. From a computational perspective there is thus little incentive to keep the history short. However, the assumptions that demand was constant over the history – and in practice data storage requirements – still call for a reasonable compromise when choosing the history length.

6.4.2. Forecast Initialization

In simulations, as well as in real life, forecasting methods have to be initialized in some way. However, we are mainly interested in the long-term behavior of an estimation or forecasting method, after any initialization effects have vanished. The standard method is to use a burn-in phase, where the simulation is executed for a number of runs until a steady state is reached. Only after that burn-in phase statistics are collected. The simulation runs during the burn-in phase are therefore ”wasted” computation time, and it should be a goal to keep the required length of the burn-in phase to a minimum.

To accomplish this, we initialize the forecast with a given mean squared error, by starting with the perfect forecast (i.e. the real demand) and adding an error term to it. The solution to equation 4.20 from section 4.4 lets us approximate the steady state Posterior Cramér-Rao Bound I_{∞}^{-1}, which we use as the initial mean squared forecast error. R This is only an approximation of the true steady state for two reasons. First, the actual mean squared forecast error will be larger than the PCRB, this is especially true for the baseline methods. Second, equation 4.20 assumes a constant measurement information matrix $M_t = M$, which is not the case during actual simulation. We use a weighted average of all nested availability situations, to compute an approximate M to use in equation 4.20.

Preliminary simulation runs showed that the approximation is quite good. Thus, with this method, simulations start much closer to the desired steady state, and therefore a relatively short burn-in phase of 50 simulation runs seems adequate. To solve equation 4.20 in REMATE, we implemented the method of Laub (1979) by porting the LAPACK (Anderson et al. 1999) routine STREXC and its dependencies to the Java language, which lets us perform the necessary re-ordering of eigenvalues in a Schur decomposition.

6.5. Simulation Results

This section presents the first simulation results. Summarizing the previous sections, there are 3 demand volumes, 3 traffic types, 10 estimation method variants and 10 independent demand realizations with 100 runs each, for a total of 900 simulations and 90000 simulation runs. Total running time was about 100 hours on a laptop computer with an Intel Core i5 2.6 GHz processor and 4GB of RAM.

We analyze performance in term of both forecast quality and achieved revenue. Forecast quality is measured with the so-called estimator efficiency, defined in equation 4.21. Revenue results are reported as a revenue loss in percent, compared to the revenue that is achievable using the real demand parameters as a forecast, the true parameter revenue.

Abbreviation	PFa no AE	PFa with AE	PFb no AE	PFb with AE	PFc no AE	PFc with AE	UKF no AE	UKF with AE
Filter	PF						UKF	
Importance Func.	Simple				Approx. Opt.		-	
Particle Count	1000		10000		1000		-	
Active Estimation	No	Yes	No	Yes	No	Yes	No	Yes

Table 6.1.: Overview of the eight different filtering method variants evaluated in this section.

6.5.1. Comparison of Filter Design Parameters

As with most heuristics, not all design decision of the filtering methods can be made from purely theoretical considerations. Therefore, we first compare six variants of the Particle Filter (PF) and two versions of the Unscented Kalman Filter (UKF) to each other. From this, we select the best PF variant and the best UKF variant and evaluate them against our benchmark methods in section 6.5.2.

Filter Variants

As noted in section 6.2, we consider two choices for the importance sampling function: an approximation to the optimal importance sampling function and a lower quality, but simple to compute one. To keep computation time to a manageable and realistic amount, we had to limit the number of particles to 1000 for the approximately optimal importance function. For the simpler variant, we could raise that limit to 10000, but also included a version with only 1000 particles in order to separate the effects of importance function and particle count. Finally, for all these and the UKF, we have a variant with and without active estimation[3]. Table 6.1 summarizes this paragraph.

Forecast Quality

Figure 6.1 compares estimator efficiency for the eight variants. For each data point, the estimated mean and approximate 95% confidence intervals are given[4] The figure shows aggregated data over all demand levels and traffic type scenarios, since we did not find

[3]see section 6.3

[4]Here the data points are in fact ratios of two experiments. We use the R-package `pairwiseCI` (Froemke et al. 2012), which implements a method proposed by Ogawa (1983), to compute the (approximate) confidence intervals.

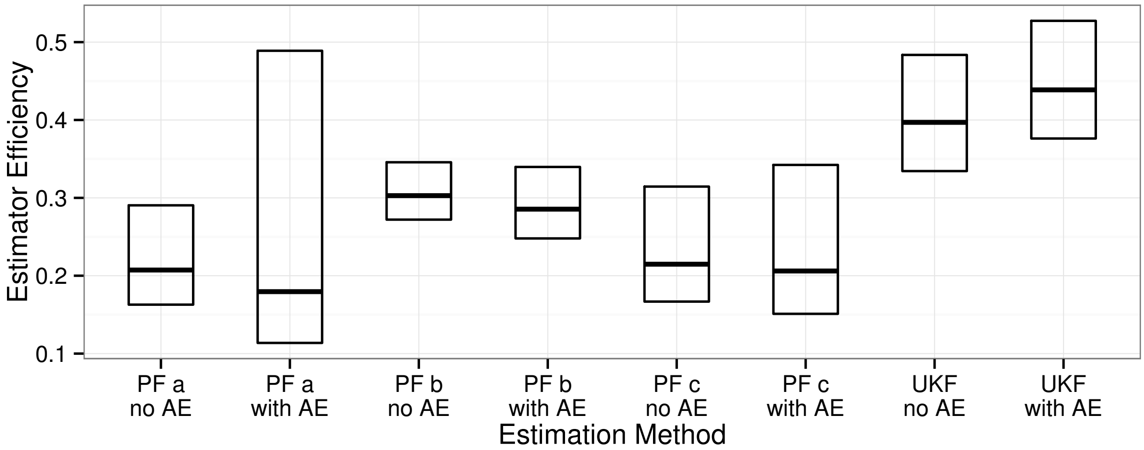

Figure 6.1.: Estimator efficiency $E = {}^{tr(\mathrm{PCRB})}/_{tr(\mathrm{MSE})}$ for filter variants, aggregated over all scenarios; the boxes represent the mean value and its 95% confidence interval; see table 6.1 for an overview over the estimation methods in this figure.

any large qualitative differences between them. The most pronounced difference exists between the UKF variants and the PF variants, with the UKF being the more efficient estimator, in many cases with statistical significance.

Among the respective filter variants, however, differences are much more subtle and do not rise to statistical significance in our study. Nevertheless, we observe a few trends: Moving from the simple importance function to the approximately optimal one (PF a to PF c), does not seem to have any measurable effect. Increasing the number of particles by a factor of 10 (PF a to PF b), however, does have a noticeable positive effect, which is on the border to statistical significance.

Active Estimation has little effect on estimator efficiency in general, with a slight negative effect for all PF variants and a small positive effect for the UKF. The expected result is an increase in estimator efficiency in all cases, since the estimators should benefit from increased variability in availabilities. This improved forecast accuracy could then offset any negative revenue effect resulting from randomizing availabilities to some extent. Our simulation study shows, however, that the positive effect on forecast accuracy is very small for the UKF and non-existent for all three PF variants. We therefore do not expect a positive revenue effect from Active Estimation.

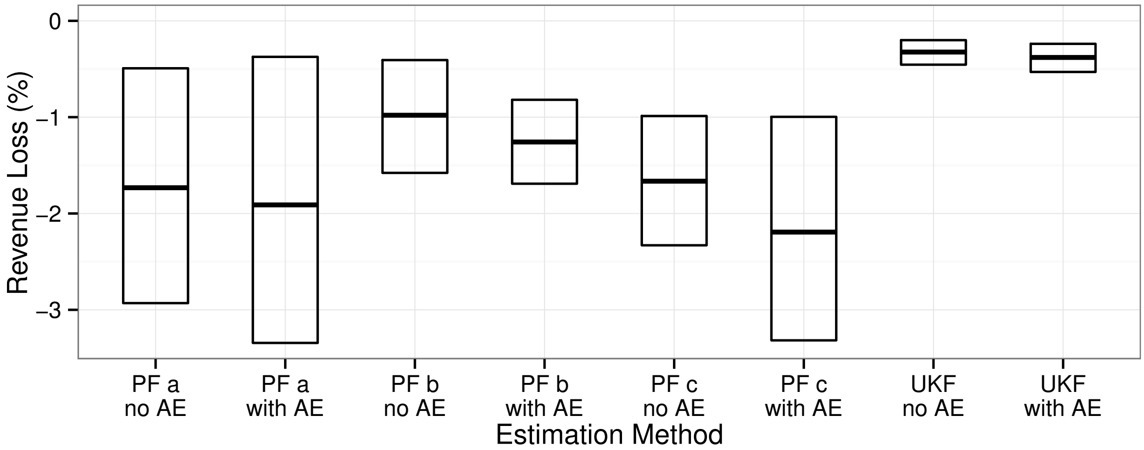

Figure 6.2.: Revenue loss compared to true parameter revenue for filter variants, aggregated over all scenarios; the boxes represent the mean value and its 95% confidence interval; see table 6.1 for an overview over the estimation methods in this figure.

Revenue

Figure 6.2 shows revenue loss compared to the true parameter revenue. Again the UKF variants perform significantly better than all PF variants, in most cases with statistical significance. As is the case for estimator efficiency, there is no perceivable difference between the two importance functions for the PF (PF a to PF c), and we observe a positive, yet not statistically significant effect when the number of particles is increased tenfold (PF a to PF b). Active Estimation has a slightly negative impact on revenue in all cases, which is the expected outcome given the results on estimator efficiency.

Conclusion

Among the PF variants, we select the simple importance function with 10000 particles (PF b) without Active Estimation. It has the best performance in both indicators, while being the conceptually simpler method at the same time. For the UKF, the decision is not as clear-cut, since the results of not agree between the two indicators. Here, we opt for the UKF without Active Estimation, since Active Estimation adds complexity to the method without showing a clear advantage. Going forward these two selected variants will simply be called PF and UKF, respectively.

Our results on Active Estimation mirror those of Aviv & Pazgal (2005), while opposing the conclusion of Carvalho & Puterman (2004), albeit in a slightly different setting.

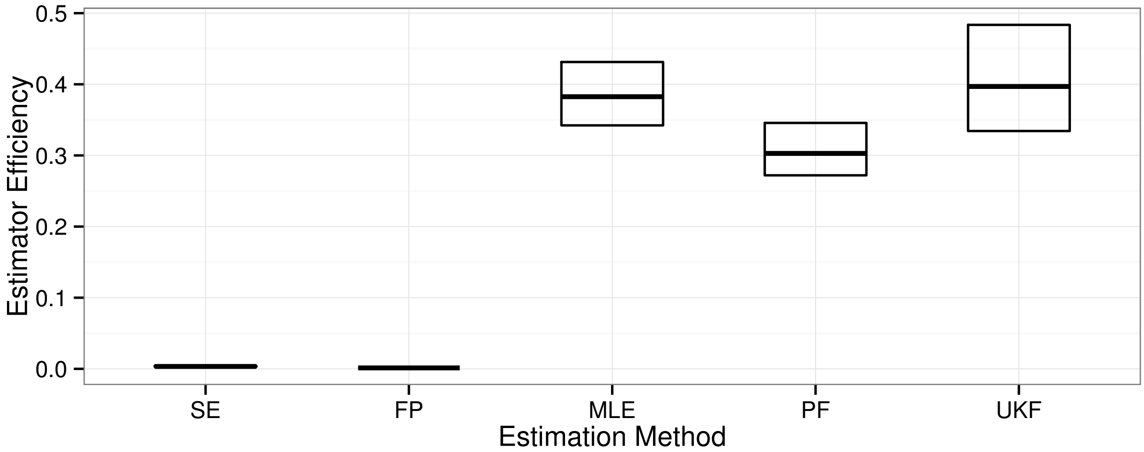

Figure 6.3.: Estimator efficiency $E = {}^{tr(\mathrm{PCRB})}/_{tr(\mathrm{MSE})}$, aggregated over all scenarios; the boxes represent the mean value and its 95% confidence interval; SE = Simple Estimation, FP = Forecast Prediction, MLE = Maximum-Likelihood, UKF = Unscented Kalman Filter, PF = Particle Filter

Our Active Estimation procedure was not able to provide additional revenue, despite improving forecast error in some cases. Since we used a simple heuristic, however, this does not imply that our estimation problem cannot benefit from active learning per se: a more refined method for active learning might lead to increased revenues after all.

6.5.2. Comparison to Benchmark Methods

In this section, we compare the selected PF and UKF variants from the previous section to our benchmarking methods. Except for the selected estimation methods, simulation setup is identical to the previous section, such that the result sets are comparable to each other. The split in two distinct results sets is merely to facilitate analysis.

Forecast Quality

Figures 6.3, 6.4 and 6.5 show the estimator efficiency obtained in the simulation. Figure 6.3 provides an aggregated overview over all demand volumes and scenarios. SE and FP produce a mean squared error that is orders of magnitude higher than the PCRB and thus efficiency is close to 0. The PF's mean squared error is about 3 times as high as the PCRB, which leads to an estimator efficiency that is lower than that of the UKF and MLE; in both cases the result is just below the border to statistical significance (the confidence intervals overlap slightly). The UKF is slightly more efficient than MLE, but

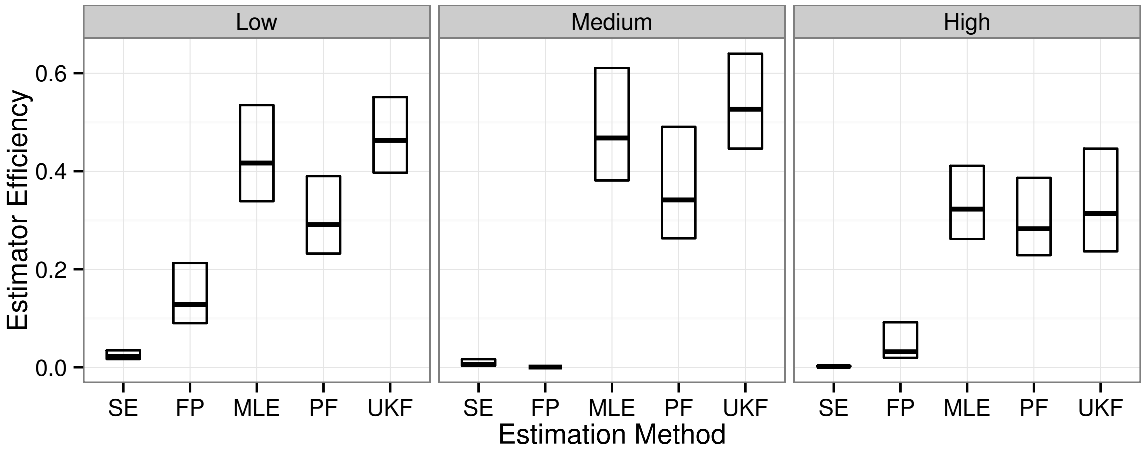

Figure 6.4.: Estimator efficiency $E = {}^{tr(\mathrm{PCRB})}/_{tr(\mathrm{MSE})}$ by demand volume; the boxes represent the mean value and its 95% confidence interval; SE = Simple Estimation, FP = Forecast Prediction, MLE = Maximum-Likelihood, UKF = Unscented Kalman Filter, PF = Particle Filter

this result is far from significant.

Figure 6.4 splits the data by demand volume. FP performs worst in the medium demand case, for low and high demand it yields a significantly lower forecast error than SE. Both filter methods suffer from a loss of efficiency in the high demand case. That loss is least pronounced for the PF, but even in the high demand case its efficiency is still slightly lower than that of the UKF and MLE.

A possible explanation is quality of availability information which degrades as the demand volume increases. If bid prices are constant zero, then availability is solely determined by fare transformation. These availabilities will thus be constant over a time period, providing perfect availability information to the estimator. If bid-prices are positive, availability may change each time a booking occurs and every time the bid price vector gets updated (once per day in the simulation). The estimators, however, are not aware of these within-time-period availability changes, they only get a rough approximation of the total amount of time a class was available during a time period, based on a linear interpolation of the bid prices at the beginning and the end of the time period.

Figure 6.5 splits the data by traffic type. For intercontinental traffic all methods show reduced estimator efficiency compared to the other two scenarios, an effect that is statistically significant for MLE and the UKF, but only partially for the PF. Qualitative results are the same for all three scenarios, with the exception that the ordering of MLE

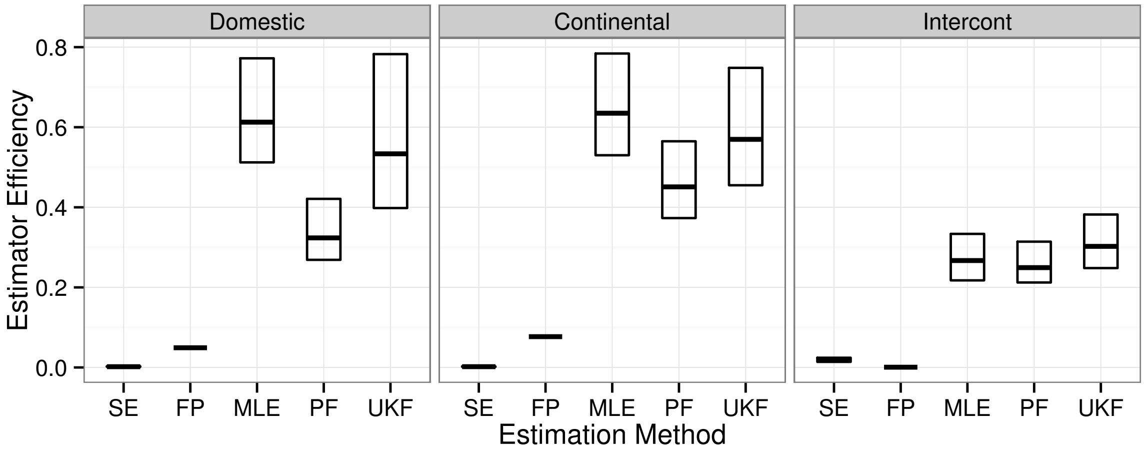

Figure 6.5.: Estimator efficiency $E = {}^{tr(\mathrm{PCRB})}/_{tr(\mathrm{MSE})}$ by traffic type; the boxes represent the mean value and its 95% confidence interval; SE = Simple Estimation, FP = Forecast Prediction, MLE = Maximum-Likelihood, UKF = Unscented Kalman Filter, PF = Particle Filter

and UKF changes between the intercontintental and the other two scenarios. In all cases, however, the difference between UKF and MLE is not statistically significant. FP showed very large variations in mean-squared error, which precluded us from computing confidence intervals in this case: since mean-squared error is in the denominator of estimator efficiency and we assume mean-squared errors to be Gaussian, too much variance causes problems by allowing a non-positive denominator with non-negligible probability. Therefore, only the means are shown for FP in figure 6.5.

Revenue

Figures 6.6 through 6.8 show the relative loss in revenue of each estimation method compared to the true parameter revenue. Figure 6.6 displays the aggregated data over all traffic types and demand volumes, while figures 6.7 and 6.8 provide more detail by splitting the data by demand volume and traffic type, respectively. Again, for each data point the mean and its 95% confidence interval are shown.

SE yields significantly less revenue than all other methods, with a total loss of about 15%, followed by FP with about 5% loss. In the aggregate over all scenarios, MLE and UKF lead to the highest revenues, with only a slight loss ($\approx 0.22\%$ and $\approx 0.32\%$) compared to the true parameter revenue, with no significant difference between them. The PF has a noticeably higher revenue loss of close to 1%, which is statistically significant compared to MLE, but just below the threshold to significance compared to the UKF.

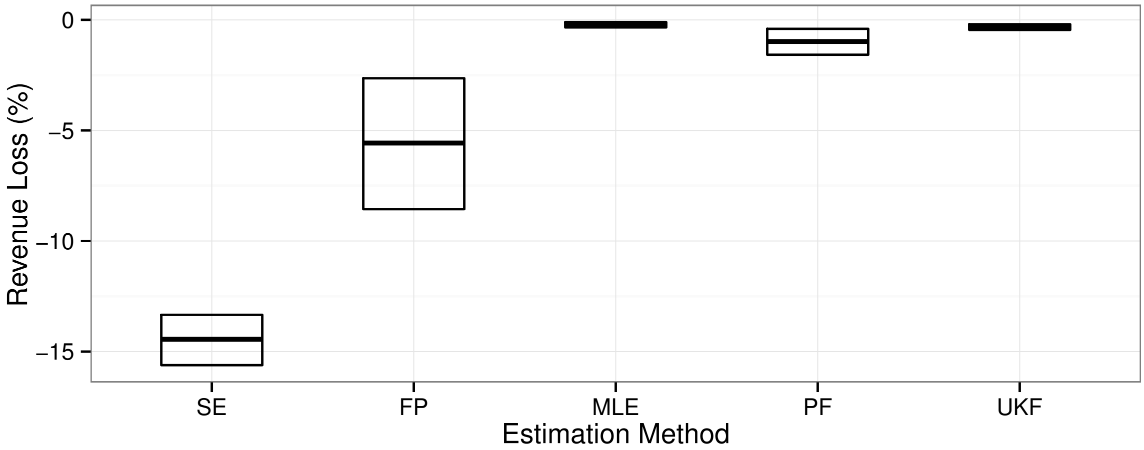

Figure 6.6.: Revenue loss in percent compared to true parameter revenue, aggregated over all scenarios; the boxes represent the mean value and its 95% confidence interval; SE = Simple Estimation, FP = Forecast Prediction, MLE = Maximum-Likelihood, UKF = Unscented Kalman Filter, PF = Particle Filter

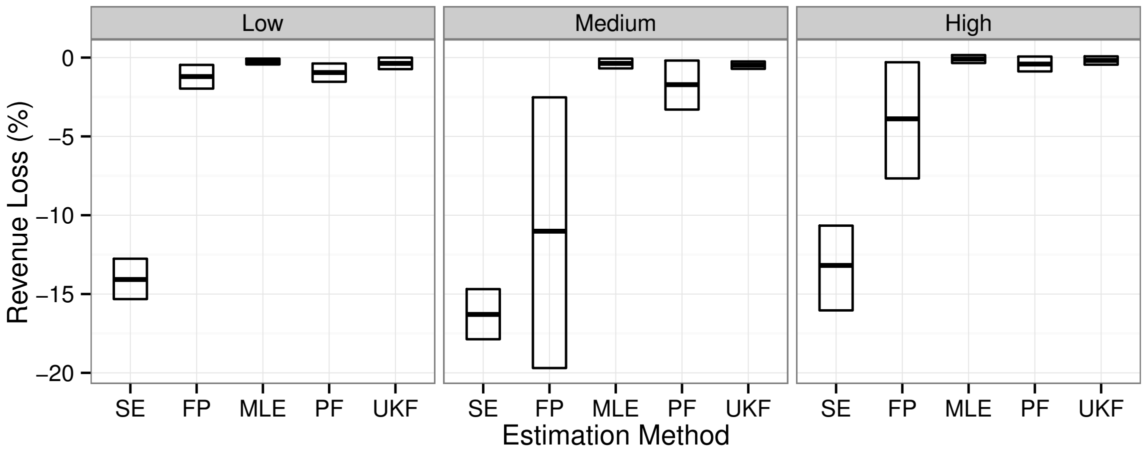

Figure 6.7.: Revenue loss in percent compared to true parameter revenue, by demand volume; the boxes represent the mean value and its 95% confidence interval; SE = Simple Estimation, FP = Forecast Prediction, MLE = Maximum-Likelihood, UKF = Unscented Kalman Filter, PF = Particle Filter

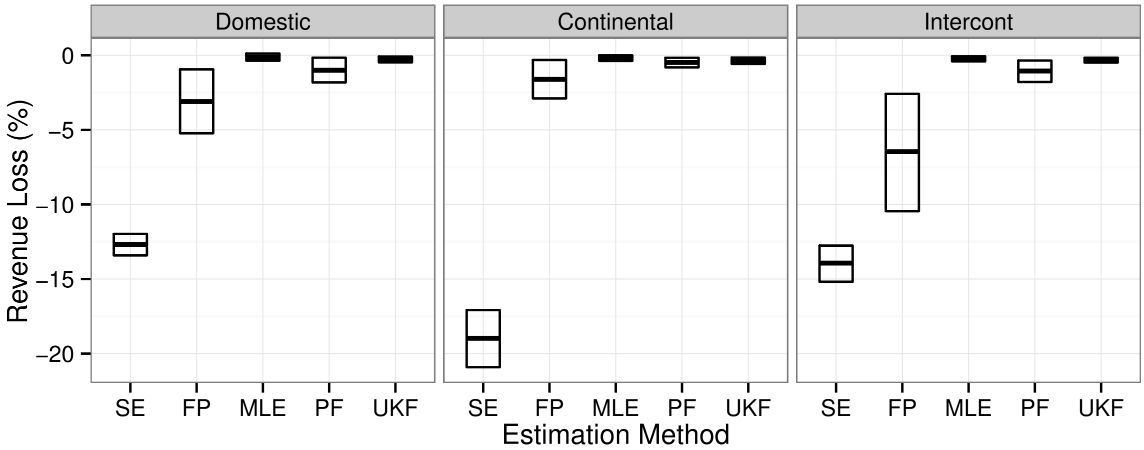

Figure 6.8.: Revenue loss in percent compared to true parameter revenue, by traffic type; the boxes represent the mean value and its 95% confidence interval; SE = Simple Estimation, FP = Forecast Prediction, MLE = Maximum-Likelihood, UKF = Unscented Kalman Filter, PF = Particle Filter

The more detailed views in figures 6.7 and 6.8 confirm that the qualitative results are the same over all demand volumes and traffic types for the revenue loss indicators. While this does not lead to any new insights, it suggests that our results and conclusions from this section are robust under various perturbations of the scenario parameters.

6.5.3. Faster Demand Change

In this section, we analyze the impact of faster demand change on the estimation methods. This is done for the domestic scenario with medium demand only. Using that setting as a baseline, we increased the rate of demand parameter change for all parameters by factors of 2 and 4. The domestic scenario was already the one with the highest rate of demand change, and the twofold and fourfold rates are well above anything we observed in our data analysis in section 5.4. From this perspective, the estimation environment of this study is therefore more challenging than anything we would expect in a real-world network. Yet, the performance of our estimation methods under these conditions may provide useful hints towards their robustness under less favorable conditions.

Figure 6.9 shows estimator efficiency for the three rates of demand change. MLE and UKF without Active Estimation loose efficiency as the rate of demand change increases. Only the UKF with Active Estimation remains roughly constant over the different change rates. This suggests that availabilities vary enough without Active Estimation to track

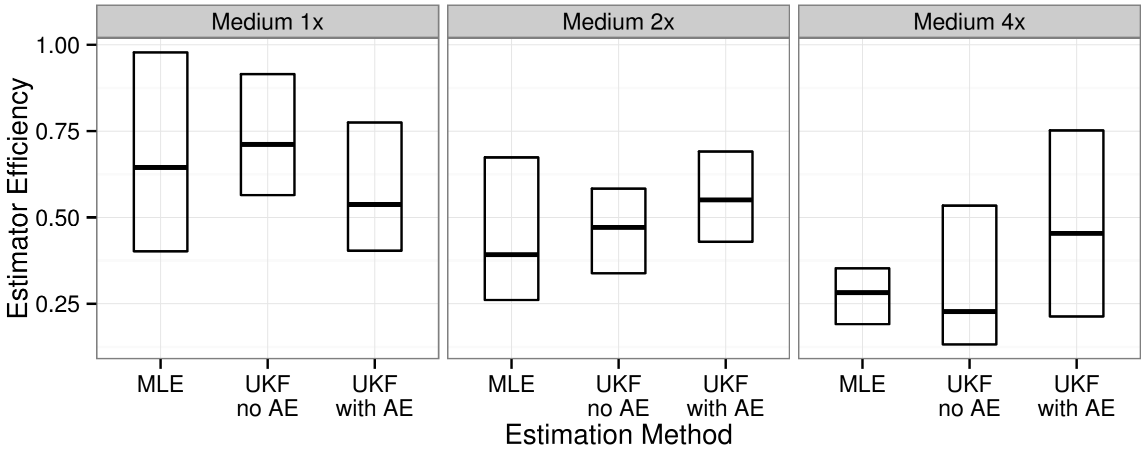

Figure 6.9.: Estimator efficiency $E = {}^{tr(\text{PCRB})}/_{tr(\text{MSE})}$ in the domestic scenario for the original rate of demand change, twice that rate and four times that rate; the boxes represent the mean value and its 95% confidence interval; MLE = Maximum-Likelihood, UKF = Unscented Kalman Filter.

the slow changing demand parameters in the original scenario, but increased variability is necessary to follow the faster changing signal in the other two scenarios.

Figure 6.10 depicts revenue loss compared to the true parameter revenue for the three rates of demand change. Revenue loss increases for all methods as the rate of demand change grows. This is the expected result, since the PCRB will grow with increasingly fast changing demand. So even if estimator efficiency is constant, mean-squared error will increase and thus more revenue loss is to be expected. The relative performance of both UKF methods improves slightly compared to MLE, however the effect is very small and not statistically significant. Again, the positive impact of Active Estimation on estimator efficiency does not translate to reduced revenue loss.

6.6. Discussion of Simulation Results

The simulation results in this chapter show that both proposed filtering methods perform generally well over a range of scenario parameters. Their estimation errors are at least an order of magnitude smaller and achieved revenue is significantly higher when compared to the SE and FP benchmark methods. Between the UKF and the PF, there is a clear advantage for the UKF in terms of both revenue and estimator efficiency. While individual results were not always statistically significant, the fact that the same pattern was present in all our different simulation scenarios makes us confident that this is the

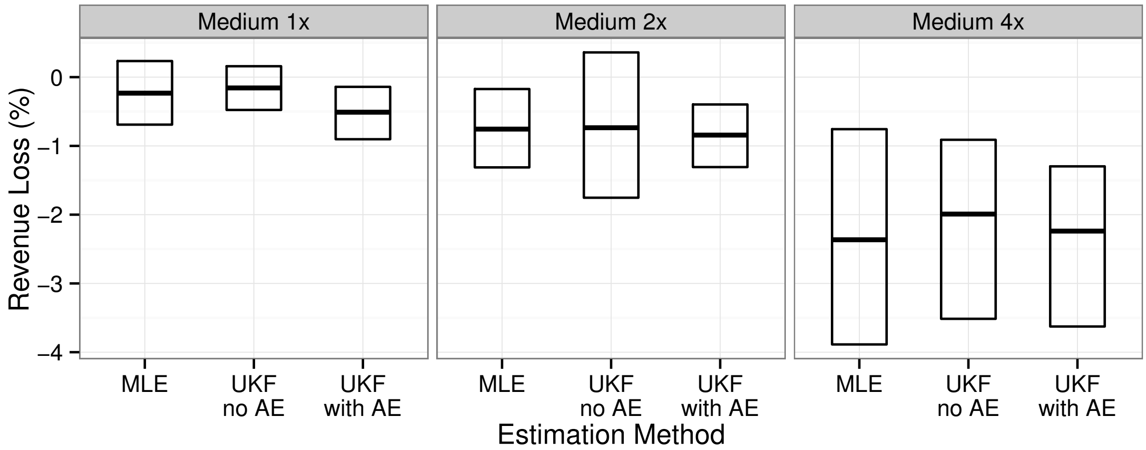

Figure 6.10.: Revenue loss in percent compared to true parameter revenue in the domestic scenario for the original rate of demand change, twice that rate and four times that rate; the boxes represent the mean value and its 95% confidence interval; MLE = Maximum-Likelihood, UKF = Unscented Kalman Filter.

correct conclusion. Comparing the UKF to MLE, however, leads to more ambiguous results. The UKF has a slight advantage over MLE in terms of estimator efficiency, but a small disadvantage in revenue loss. However, both of these effects are far from statistical significance and we consider the performance of UKF and MLE in this simulation study to be equivalent.

Thus, the results of this chapter clearly suggest the use of either MLE or the UKF for demand estimation. Yet, no recommendation for choosing between those two methods can be given from the simulation results alone. However, the UKF has one additional benefit: It provides an approximation of the estimator's uncertainty in the form of the covariance matrix P_t[5]. This additional information can be useful in several contexts. Estimator uncertainty can for example be accounted for during optimization, can inform a human analyst about the confidence in the system forecast she should have, and can also be used to dynamically adapt the level of detail for forecasting. This latter option will be explored further in chapter 8.

Given the generally excellent performance of the UKF and its additional utility over the MLE, the UKF is our estimation method of choice for the simulation studies in subsequent chapters. The general discussion however will remain independent of any particular estimation method. Thus, if it turns out that for a different demand model, say, the PF is a better choice, the methods and theoretical results in the remainder of

[5] see section 6.1

this thesis still apply.

Part III.

The Problem of Small Numbers

7. Quantifying the Problem of Small Numbers

The simulation results presented in chapter 6 promise estimation and revenue performance close to their theoretical upper bounds – the PCRB and true parameter revenue, respectively. This chapter will show that the results presented so far come from a very favorable scenario where all demand is concentrated on a single origin-destination pair (O&D). In real airline networks, demand is distributed over a large number of such O&Ds and it is considered appropriate to estimate demand for each of those origin-destination pairs individually in practice (Boyd & Kallesen 2004). Moreover the distribution of demand over O&Ds is highly non-uniform, varying over several orders of magnitude, with some very large markets and many extremely small markets.

As briefly argued in section 2.2, the problem of small numbers is different from the problem of small sample size. When demand estimation is done at a finer level of detail, and sales data is collected at a finer level as well, this actually increases sample size. Even if no bookings occur for a given combination of product, price point and period of time this still constitutes an observation or sample. Such an observation also carries meaningful information, namely that there was probably little demand for the given product, at the given price point, in the observed period of time. One might be inclined to believe that a zero booking observation provides information on overall demand volume, but that no additional information on customer choice behavior can be gained without an actual booking event. Yet, this is not the case either. Say, an airline has offered a product at a low price for some time in the past and observed positive demand for it. Now, it stops offering the low price point, and instead only a much higher price is available. If we now observe zero bookings instead of some positive number, this provides information on the willingness-to-pay of the airline's customers, namely that it is lower than the new high price point. Therefore, observations of zero bookings are not necessarily inferior to observations of a positive number of bookings.

Nevertheless, we show that estimating the parameters of a choice model is fundamentally more difficult for small volume markets than for large volume markets. While estimation error may still remain close to its theoretical bound, this bound becomes arbitrarily large for small volume markets. A simulation study shows further that this results in large revenue losses compared to the true parameter revenue.

7.1. A Property of the Posterior Cramér-Rao Bound

The PCRB, introduced in section 4.4, gives rise to a structural result that we call the "Problem of Small Numbers". It arises in practical implementations of revenue management methods: Increasingly sophisticated revenue management optimizers require greater amounts of input data on finer detail levels (Bartke et al. 2013). This input data usually has to be estimated from booking data and projected into the future. There is a common, yet often vague understanding among practitioners that the quality of this input data will decrease as more and more parameters have to be estimated from the available booking data. This decrease in quality will at some point offset any additional benefit from a more sophisticated revenue management optimizer.

In our particular framework, we can prove the first part of the assertion, namely that the mean-squared error of choice parameter estimates will tend to infinity as the number of bookings observed per data point goes to zero. The subsequent interplay with the optimizer is far from trivial, and will not be treated rigorously in this work. However, it seems reasonable to assume that an arbitrarily large variance in the input parameters can lead to highly sub-optimal optimization results.

Theorem 1 *Let the choice function $H_a(x)$ be parameterized such that it can be written as the product of an overall arrival rate λ and a choice probability function $C_a(x)$:*

$$H_a(\lambda, x) = \lambda \cdot C_a(x),$$

where $0 \leq \sum_i c_{a,i}(x) \leq 1$, $\lambda > 0$ and all parameters are identifiable from the set of available observations. Formally, the latter condition is true if and only if the matrix J with rows

$$\begin{pmatrix} c_{a,i}(x_\tau) & (\nabla c_{a,i}(x_\tau))^T \end{pmatrix} \qquad \text{for } \tau \leq t \text{ and } i : c_{a,i}(x_\tau) > 0$$

has full column rank. Further, assume that there is no initial information[1]: $I_0 = 0$.

[1] The case when $I_0 > 0$ holds similarly if we let $t \to \infty$. Intuitively this is the case since any initial

Then, for any estimator g of the choice parameters x (excluding λ), the mean-squared error will become arbitrarily large as the arrival rate λ tends to zero:

$$E[(g(B_t) - x_t)(g(B_t) - x_t)^T] \to \infty \qquad \text{for } \lambda \to 0. \qquad \square$$

PROOF Inserting the choice function into equation 4.19 yields the measurement information matrix

$$M_t = E \begin{pmatrix} \frac{1}{\lambda_t} \sum_i c_{a,i}(x_t) & \sum_i (\nabla c_{a,i}(x_t))^T \\ \sum_i \nabla c_{a,i}(x_t) & \lambda_t \sum_i \frac{(\nabla x_t c_{a,i}(x_t))(\nabla x_t c_{a,i}(x_t))^T}{c_{a,i}(x_t)} \end{pmatrix} \tag{7.1}$$

The sum of the measurement information matrices up to time T is an upper bound to the information matrix at time T, because the influence of time evolution can only decrease the information on hand. In formulas, since $I_t \leq (I_t^{-1} + Q)^{-1}$ (using $Q \geq 0$, since it is a covariance matrix), we have that $I_T \leq \sum_{t=1}^{T} M_t$. We will use this to compute a lower bound on the estimator error: $I_T^{-1} \geq \left(\sum_{t=1}^{T} M_t\right)^{-1}$.

Define $\lambda_t = \lambda \cdot s_t$, the matrix J' with rows

$$\begin{pmatrix} s_\tau \cdot c_{a,i}(x_\tau) & \lambda \cdot s_\tau \cdot (\nabla c_{a,i}(x_\tau))^T \end{pmatrix} \qquad \text{for } \tau \leq t \text{ and } i : c_{a,i}(x_\tau) > 0$$

and the diagonal matrix Δ with diagonal entries $\frac{1}{c_{a,i}(x_\tau)}$, $\tau \leq t$ and $i : c_{a,i}(x_\tau) > 0$. Then, using equation 7.1, we find that the right-hand side has the form:

$$\begin{aligned} \left(\sum_{t=1}^{T} M_t\right)^{-1} &= (E[J'^T \cdot \Delta \cdot J'])^{-1} = \begin{pmatrix} \frac{1}{\lambda} a & b^T \\ b & \lambda D \end{pmatrix}^{-1} \\ &= \begin{pmatrix} \frac{\lambda}{a - b^T D^{-1} b} & -b^T (aD - bb^T)^{-1} \\ -(aD - bb^T)^{-1} b & \frac{a}{\lambda} (aD - bb^T)^{-1} \end{pmatrix} \end{aligned} \tag{7.2}$$

where the scalar a, the vector b and the matrix D are all independent of λ. Since J has full column rank by assumption and J' equals J except for a rescaling of its rows and columns, J' also has full column rank. This guarantees that $J'^T \cdot \Delta \cdot J'$ is invertible and furthermore strictly positive-definite. The lower right block of equation 7.2, $(aD - bb^T)^{-1}$, is therefore also positive-definite. With that, letting $\lambda \to 0$ shows that the lower right block of the PCRB will become arbitrarily large. ■

information will loose its value over time and in the limit it will be zero. We provide no formal proof however.

As stated above, this loss of estimation accuracy cannot be explained by small sample size. Instead, we try to provide some intuition based on the ratio of two Poisson processes. In order to gain insight on customer choice behavior, e.g. to find the fraction of customers willing to buy-up to a higher price point, we essentially need to estimate the ratio of two arrival rates from observed booking counts. Say we have a Poisson count with arrival rate λ in the denominator and one with arrival rate $s \cdot \lambda$ in the numerator. Now, consider the distributions of booking counts in terms of their means and standard deviations. In the denominator we have $\lambda \pm \sqrt{\lambda}$, and in the numerator $s \cdot \lambda \pm \sqrt{s \cdot \lambda}$. Hence, the ratio is

$$\frac{s \cdot \lambda \pm \sqrt{s \cdot \lambda}}{\lambda \pm \sqrt{\lambda}} = \frac{s \pm \sqrt{\frac{s}{\lambda}}}{1 \pm \sqrt{\frac{1}{\lambda}}}. \tag{7.3}$$

For constant s and decreasing λ the standard deviations in both the numerator and denominator start to dominate the equation. As such, estimating the ratio of the two arrival rates $\frac{s \cdot \lambda}{\lambda} = s$ becomes increasingly more error-prone. More precisely, Price & Bonett (2000) examine methods for constructing (approximate) confidence intervals for the ratio of two Poisson rates. The size of these intervals grows when both observed counts become smaller proportionally, which is particularly easy to see in their method based on the Poisson log-linear model. Accordingly, the problem of small numbers in revenue management is a consequence of the combination of:

1. Estimating the ratios of arrival rates
2. Increasing relative standard deviation of Poisson counts as arrival rate becomes small

Again, the first point seems to be inevitable when estimating customer choice models. Concerning the second point, even the discrete distribution with the smallest variance for an arrival rate $0 < \lambda < 1$, the Bernoulli distribution, has a variance of $\lambda - \lambda^2$, which tends to λ, the variance of the Poisson process, as λ tends to zero. Moreover, practitioners mostly believe that the Poisson assumption errs on the side of too little variance, rather than too much (Walczak 2006).

Finally, note that the choice probability function $C_a(x)$ may represent a customer's choice between multiple products, but also other modulations of the overall arrival rate, such as seasonal or weekly patterns.[2] Hence, estimating the parameters of any of such effect becomes arbitrarily error-prone as the arrival rate tends to zero.

[2]This was pointed out to us by Karl Isler of SWISS Airlines.

7.2. Simulation Study

In this section, we confirm the theoretical results on forecast error from the previous section and investigate the revenue implications.

7.2.1. Simulation Setup

Basic scenario and demand setup remains as described in chapter 5. The scenarios have to be expanded however, to include multiple O&Ds. The focus remains on a single leg with capacity 100, but now there are also 29 feeder flights for this leg. Thus, there is one local O&D and 29 transfer O&Ds for a total of 30 O&Ds. The number 30 was chosen such that the simulation remains computationally tractable while still providing enough opportunities for forecast merging. In practice, there are many more O&Ds traversing a single leg (more than 500 in our data set), however many of them are so small that the probability of having a passenger from that O&D on a particular flight is almost zero. The capacities of the feeder flights are large enough that they have no limiting influence during optimization and can therefore be considered infinite for all practical purposes.

We analyze two distinct scenarios: "continental" and "intercontinental" where the trunk legs are continental or intercontinental flights, respectively. Note that the continental scenario therefore contains both continental and intercontinental traffic, while the intercontinental scenario only contains intercontinental traffic. Therefore, the band of price elasticities is expected to be wider in the continental than the intercontinental case.

The UKF is exclusively used to estimate demand according to the conclusions from chapter 6.

Passenger Volume Distribution

To find the variation in price-sensitivity and passenger volume over the O&Ds, we analyzed real-world O&D data from two outbound flight legs for the complete year of 2012, a continental and an intercontinental leg. We found a total of 536 and 961 O&Ds traversing those two legs, respectively. Figure 7.1 shows the distribution of passenger volume over these O&Ds, excluding the local O&D which accounts for roughly two thirds of all passengers on the continental leg and one third on the intercontinental leg. Still, even without local traffic, the distribution of passenger volumes is highly uneven, with few large O&Ds and many small O&Ds. In fact, on the majority of O&Ds there is on

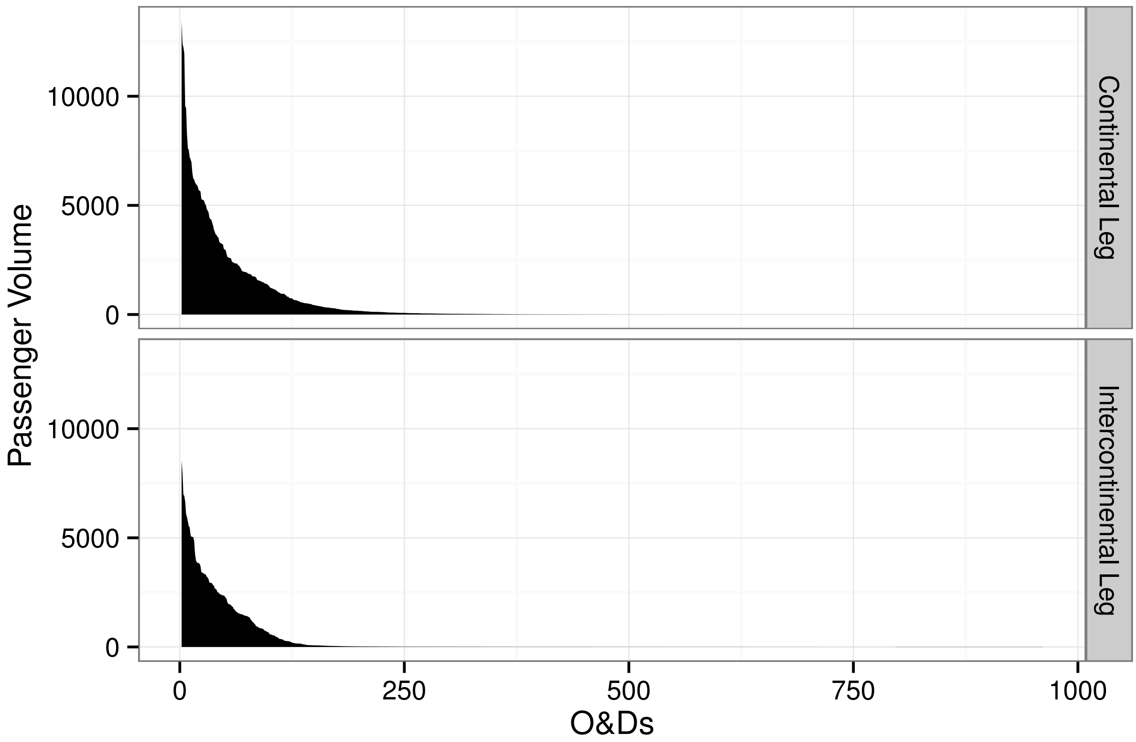

Figure 7.1.: Distribution of passenger volume over the O&Ds traversing a single leg, by type of leg, excluding the local O&D; data from 2012.

average less than one booking per departure, underlining the prevalence of the issue of small numbers for a typical airline network.

In the simulation scenario, we keep the share of local traffic exactly as in the data and then sample 29 O&Ds from the total set of O&Ds and use their relative passenger volumes. The sampling is deterministic, selecting every sixth O&D until the total number of 29 is reached. The smallest O&D in the simulation then has a booking probability of about 1 in 1000, such that we expect about one booking on that O&D over all 10 replications of the 100 simulation runs. Consequently, selecting even smaller O&Ds in addition to the present ones should only have a negligible impact on our simulation results. Nevertheless, our sampling of O&Ds tends to exaggerate O&D size in our simulation, such that small number effects in real life could be even more pronounced than in our simulation.

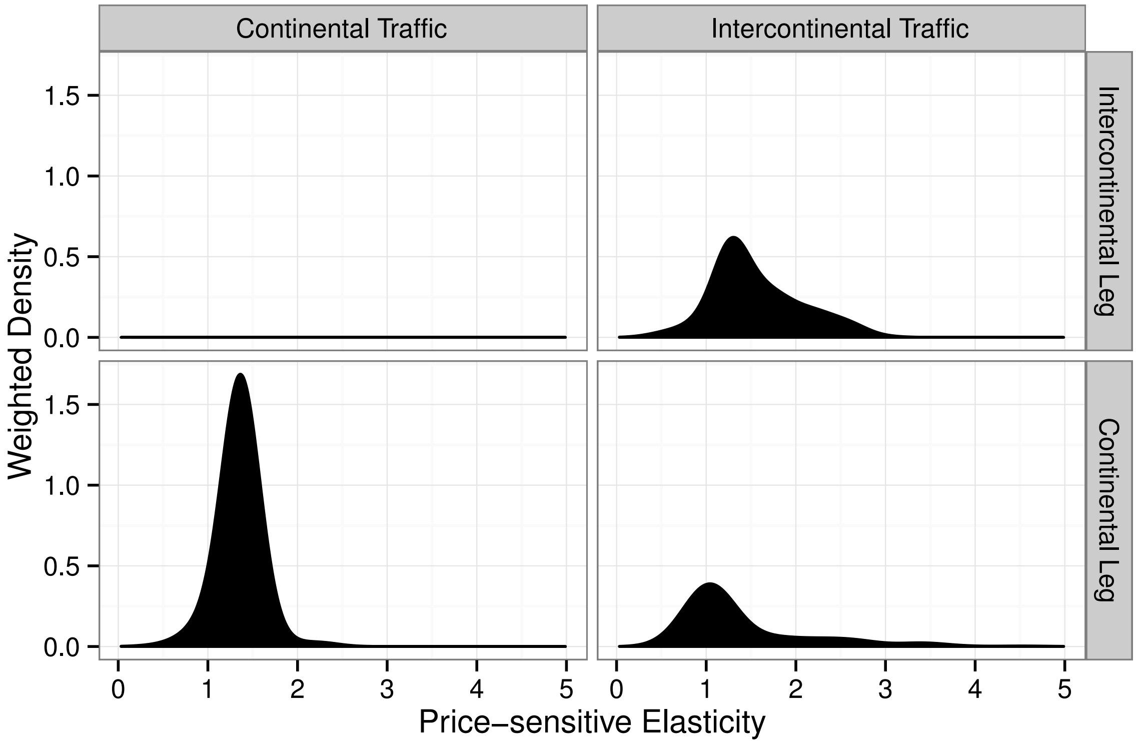

Figure 7.2.: Distribution of approximate price-sensitivity parameter over the O&Ds traversing a single continental leg, split by O&D traffic type and leg type; the figure shows a density estimate of the original data with each O&D weighted by its passenger volume; data from 2012.

Price-sensitivity Distribution

The available data set also includes average revenue per passenger for each O&D. Using the same method as in section 5.4, we computed approximate price-sensitivity parameters from that. The distribution of these, shown as a density estimate weighted by passenger volume, is plotted in figure 7.2. Since a different base price is used for intercontinental traffic, the data has been split in intercontinental traffic and in domestic and continental traffic. While the peak of the distribution is roughly Gaussian, it has a fat right-tail, which is especially pronounced for the intercontinental case.

The available data does not include the booking horizon dimension. We therefore compute a price-sensitivity factor for each O&D, which is the quotient of the O&Ds price-sensitivity parameter over the average of all O&Ds. The same deterministic sampling of O&Ds as above is used to determine price-sensitivity factors for each of the 30 O&Ds in the simulation. These factors are then applied to the price-elasticity curve from section

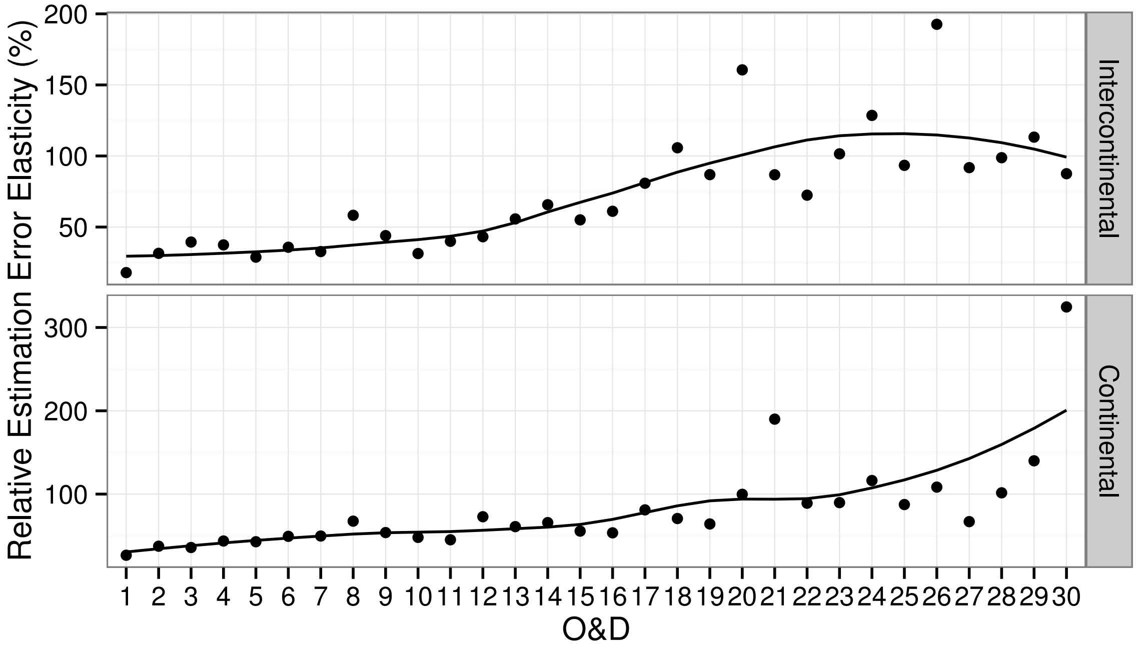

Figure 7.3.: Relative estimation error $\frac{\sqrt{MSE(x)}}{E(x)}$ of price elasticity parameters over O&Ds and by type of trunk leg; O&Ds are sorted by size in descending order; data points represent the mean of ten independent trials; smoothed line is a visual aid.

5.4 to yield the final price-sensitivity parameters for each O&D in the simulation.

Both, the volume and the elasticity parameters defined here are merely initial values for the first departure. Subsequent values are again the realizations of an AR(1) process as described in section 4.3 with the parameters from section 5.4.

7.2.2. Simulation Results

Figure 7.3 shows the relative estimation error of the price elasticity parameters over the 30 O&Ds and by type of trunk leg. The data points represent the mean of ten independent trials and the smooth line has been added as a visual aid. We define relative estimation error as the square-root of mean-squared error divided by the mean of the parameter. The measure is as such related to the relative standard deviation and is invariant under scale or unit transformations. Estimator efficiency which was used in the previous chapter to gauge estimation performance is inadequate for the purposes of this study: the effect of small numbers can not be seen in the estimator efficiency measure (see equation 4.21) since numerator and denominator are influenced in the same

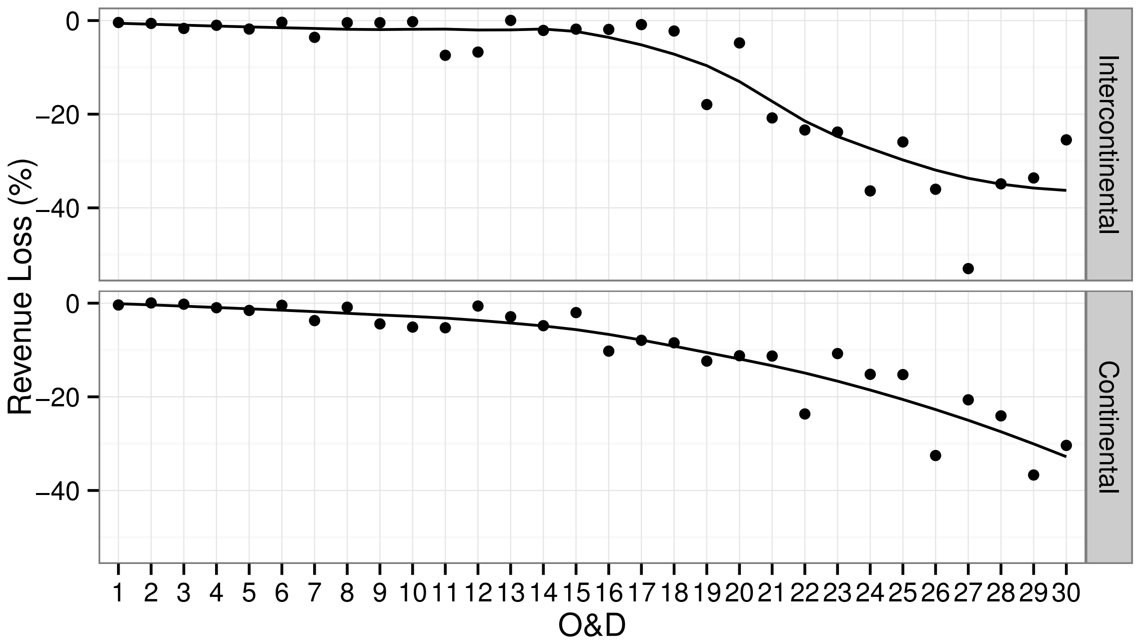

Figure 7.4.: Percentage revenue loss compared to true parameter revenue over O&Ds and by type of trunk leg; O&Ds are sorted by size in descending order: data points represent the mean of ten independent trials: smoothed line is a visual aid.

way.

Figure 7.3 confirms the prediction from theorem 1 that estimation error rises significantly with smaller O&D size. The relative estimation error is about 5 and 12 times larger for the smallest O&D compared to the largest O&D for the intercontinental and continental trunk legs, respectively. Moreover, with estimation errors over 100%, achieving optimal revenue results seems highly unlikely.

This manifests itself in figure 7.4 which depicts relative revenue loss compared to the true parameter revenue over O&Ds. The relative gap between the simulated revenue and the true parameter revenue becomes larger with decreasing O&D size. For both scenarios, revenue loss is around 30% for the smallest O&Ds, which represents a very significant revenue loss in the practice of revenue management. Since this loss is concentrated on the smallest O&Ds however, the overall revenue loss is much smaller. It totals to $1.96 \pm 0.34\%$ and $1.89 \pm 0.39\%$ over all O&Ds for the intercontinental and continental scenarios, respectively[3]. Given that the revenue loss in the single O&D scenarios of chapter 6 was much smaller at roughly 0.3% for the UKF, the effect of small numbers

[3]The ranges mark the bounds of the 95% confidence interval.

causes a loss of revenue of more than 1.5% in our simulation study.

7.3. Chapter Conclusion

The practitioners' intuition on the problem of small numbers is clearly supported by theorem 1 and our simulation results. Theorem 1 suggests that this issue arises for customer choice parameters, in particular. So, while O&D-based revenue management systems with independent demand face data sparsity as well, the negative impact of small numbers really only arises with the introduction of choice-based demand models. It has been observed that the adoption of choice-based revenue management systems has been reluctant, even though they have been discussed in the research community for numerous years and its revenue potential has been shown in multiple simulation studies (Weatherford & Ratliff 2010). The problem of small numbers may well be one of the reasons for this.

Unfortunately, theorem 1 tells us that the problem of small numbers cannot be solved by simply finding a better estimation method. Strictly speaking, the problem of small numbers cannot be *solved* at all, but it can be, at least partially, avoided. In the following chapter, we propose a forecast merging or clustering method that increases the passenger volume per elasticity estimate by merging the information from multiple O&Ds. We will show that this can mitigate the problem of small numbers to a great extent.

8. Forecast Merging

In chapter 7 on the problem of small numbers, we showed that estimation errors for customer choice parameters become arbitrarily large as the overall market volume decreases. Moreover, a simulation study revealed that this reduction in forecast quality leads to a measurable loss in overall revenue. In this section, we aim to mitigate this problem by merging multiple individual estimates to form a single new estimate. The objective is to reduce the overall mean squared estimation error as far as possible and thus regain lost revenue.

8.1. Multiple Levels of Clustering and Aggregation

For typical dependent demand models, estimation will be the computationally most intensive step. If computational tractability was of no concern, no clustering or aggregation would be required: the estimation procedure would read in all booking events (Passenger Name Records) and create a demand estimate for the airline's complete network. A mixture model would help reduce the estimation error for O&Ds with low passenger volume. However, this overall estimation problem would have billions of parameters and millions of observations. Reliably finding the global optimum in the likelihood function of such a problem seems intractable. Therefore, in practice, the estimation problem is split into a large number of smaller problems and sales data is often aggregated to a certain level. We call each of these smaller estimation problems an *estimation cluster*. In our simulation study, each O&D is a separate estimation cluster, in line with airline practice (Boyd & Kallesen 2004). Figure 8.1 illustrates the structure of this approach.

By dividing the problem into smaller sub-problems, the estimation step becomes tractable, however some drawbacks come attached. First, information is lost when aggregating bookings and availability, that is when the "depth" of the data is reduced. Intuitively, when the bookings of several different availability situations get aggregated, the information of which number of bookings happened under which particular availability situation is lost. This can be illustrated with the Fisher measurement information

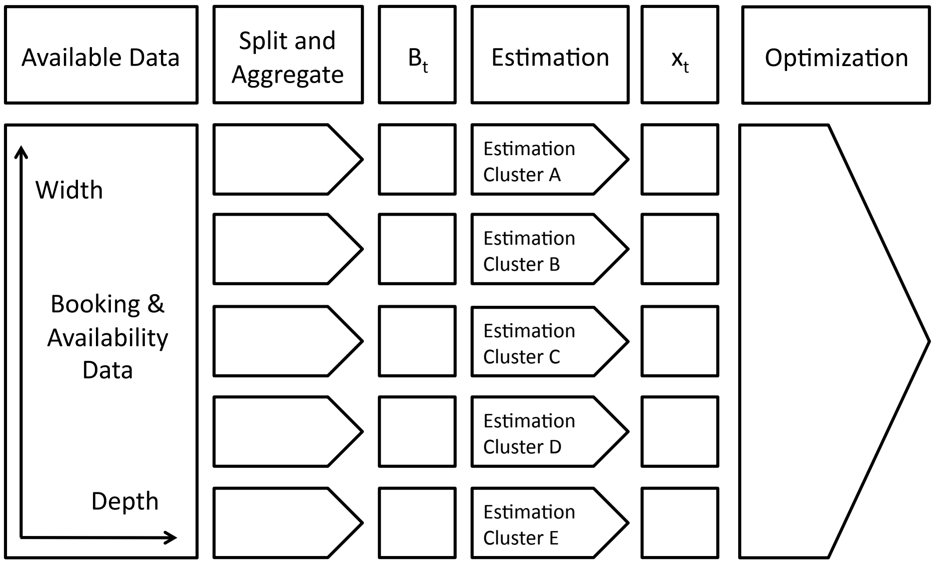

Figure 8.1.: Demand estimation in real-world applications: the global demand estimation problem is partitioned in many small and independent sub-problems to keep it computationally tractable.

matrix:

$$\begin{aligned} M_{sep} &= E\left[\sum_i \frac{(\nabla_x h_{a_i,i}(x))(\nabla_x h_{a_i,i}(x))^T}{h_{a_i,i}(x)}\right] \\ &\geq E\left[\frac{(\sum_i \nabla_x h_{a_i,i}(x))(\sum_i \nabla_x h_{a_i,i}(x))^T}{\sum_i h_{a_i,i}(x)}\right] = M_{agg} \end{aligned} \tag{8.1}$$

where M_{sep} is the measurement information matrix when bookings b_i are observed separately for each availability situation a_i and M_{agg} is the measurement information matrix when only the sum of bookings $\sum_i b_i$ over all availabilities a_i is known. Note that the act of aggregating the observations simply switches the order of summation and division/multiplication in the equation. Intuition strongly suggests that the inequality $M_{sep} \geq M_{agg}$ holds, because aggregating observations should never increase the amount of information in the data. A formal proof is given in section 10.3 of the appendix. In practice, a reasonable balance between information content, computational tractability and data storage requirement will have to be found. The properties of the Fisher measurement information matrix reported above could potentially guide such a trade off. This is however not in the scope of this work.

Instead, we focus on the second issue: there is no information exchange between estimation clusters. In a real-world airline network, many markets will have similar cus-

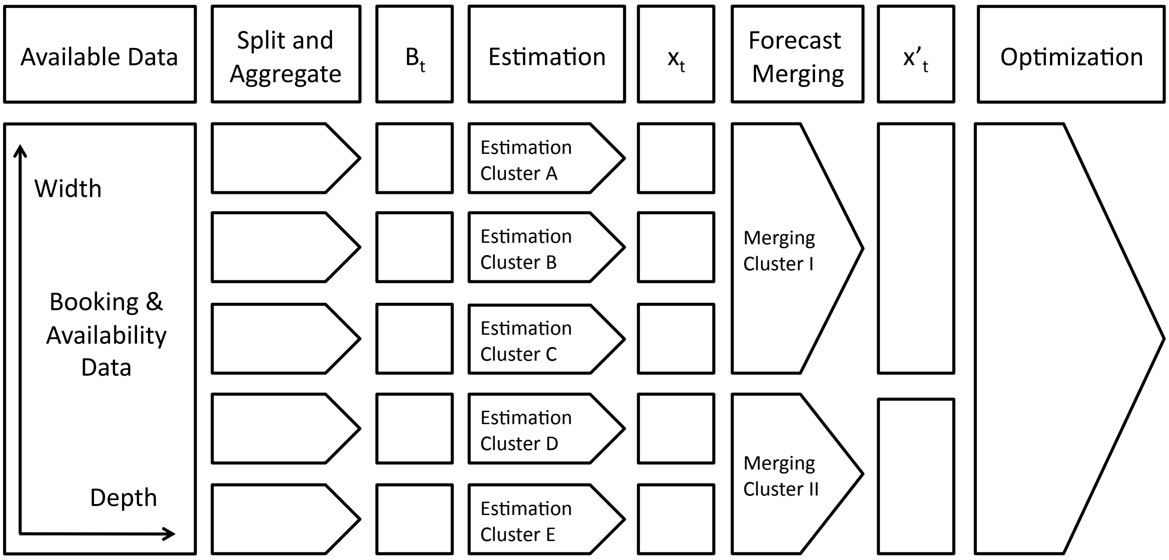

Figure 8.2.: Introduction of the proposed forecast merging step between estimation and optimization to re-enable a simple form of information exchange between estimation clusters.

tomer choice characteristics. As evident from figure 7.2, the support of the distribution of price elasticity over O&Ds is much smaller than that of O&D volumes, with most price elasticities in the same order of magnitude and many of them concentrated around the modal value. In the theoretical, complete estimation problem prior knowledge about this similarity can be used to improve the individual estimates. This ability is lost when estimation takes place individually for each estimation cluster, since each of the estimation sub-problems is assumed to be independent from the others. In the remainder of this section, we introduce a heuristic to re-enable a simple form of information exchange between estimation clusters by merging the estimates from several estimation clusters. We call a set of merged estimation clusters a ”merging cluster”. Figure 8.2 is a modified version of figure 8.1, introducing the additional forecast merging step between estimation and optimization.

8.2. Background

Combining separate forecast methods to form a new, more accurate combined forecast has been studied extensively since the seminal article by Bates & Granger (1969). D. W. Bunn (1988) and Clemen (1989) provide reviews of the then current literature, the latter being the more extensive and including more than 200 annotated references. Timmermann (2006) has surveyed the field more recently. Our work in chapter 8 is related insofar as the goal of improving forecast accuracy is identical and much of the involved equations are also the same. However, in our work, all forecasts come from the same forecasting method but have been made for separate, yet related entities. Furthermore, these forecasts are multi-variate and we wish to combine them only partially.

Our methodology is related to the article of Bates & Granger (1969). They propose a linear combination of individual forecasts to form a new, more accurate combined forecast. The weighting factors that they chose are such that the combined forecast has minimum variance. We start with a different interpretation, namely that of forming a conditional distribution, however the resulting factors are identical to those of Bates & Granger (1969). While we explicitly use the information on error variances available, we ignore their correlation. This is justified by D. Bunn & Topping (1984) who show through theoretical considerations and a simulation study that ignoring the error correlation between the combined forecasts can actually improve forecast performance when sample size is small. This is confirmed by Winkler & Makridakis (1983) whose results indicate that while variances should be considered during forecast combination, error correlations should not.

Anandalingam & Chen (1989) show that combining forecasts with a joint multivariate normal error distribution can be equivalent to Kalman filtering. Since our individual estimates already come from variants of the Kalman Filter, we can regard our overall forecasting technique, including estimation and combination, as a hierarchy of filters.

Most of the literature is focused on forecasting economic time series. Lemke & Gabrys (2008) and Lemke et al. (2013), however, propose the use of forecast combinations in airline revenue management, to forecast demand and cancellations. Again, in contrast to our work, they combine forecasts from different forecasting methods.

Both, Clemen (1989) and Lemke & Gabrys (2008) note that combined forecasts almost universally outperform single forecasts in empirical studies, even when simple combination rules - such as the simple average - are used.

Duncan et al. (1993) describe an approach to forecasting a set of univariate time-series, which are related through a hierarchical model: they assume that the parameters of the individual time-series are drawn from a common, Gaussian distribution. Duncan et al. propose a method that uses a variant of the Kalman Filter (a multi-state Kalman Filter) to estimate each individual time-series and then adjusts the individual estimates towards the overall mean. The size of the adjustment is dependent on the variance of the individual estimates and the variance between them.

Our approach differs from that of Duncan et al. (1993), since in our model the low-level time-series are already multivariate. Additionally, we do not believe that all forecasted entities are related to the same degree. Instead, we assume that some entities are known to be more closely related than others. This fact suggests the use of a clustering method, which we propose in this chapter. Within each cluster, we follow the general approach of Duncan et al. (1993), in the sense that we start with the individual estimates and then apply a relatively simple correction to them that depends on their respective variances.

8.3. Our Method for Forecast Merging

In this section, we present a method for merging estimates from different estimation clusters. Initially, the algorithm performs a hierarchical clustering of the estimation clusters to form a set of merging clusters. This hierarchical clustering is based on the expected reduction in mean-squared estimation error. Estimation clusters that offer the highest such reduction are merged first, continuing to the point where no further reduction can be expected.

Neither hierarchical clustering nor forecast merging are entirely new concepts. The novelty of this approach comes from the context to which it is applied and from combining the two. Traditionally, forecasts from different forecasting methods are combined, and in that case usually all available forecasts are used. In our case, forecasts come from the same estimation method but from a range of related, yet not identical entities. Due to that, combining all forecasts into a single one is certainly not ideal. Therefore, we add the hierarchical clustering procedure to determine which forecasts to merge and which to keep separate. While hierarchical clustering is not new in itself, we develop a new linkage criterion that determines the sequence in which clusters are formed. Since the objective is to reduce mean-squared estimation error as much as possible, this linkage criterion is defined such that those forecasts are merged first, for which the reduction in mean-squared error is highest.

As described below, computing the expected reduction in mean-squared estimation error requires the use of external data sources, such as long-term time-series and expert opinions for best results. We therefore assume that, in practice, this clustering hierarchy is determined offline and held constant for longer periods of time. To model this in the simulation, the clustering hierarchy is computed at the start of the simulation using perturbed true data and then held fixed for the remaining simulation runs. Subsequently, we use the new current demand parameter estimates and covariances and merge them in the order prescribed by the initial clustering hierarchy.

Our forecast merging algorithm requires an estimate of the demand parameters for each estimation cluster in the form of a multivariate Gaussian distribution, given by its parameters: mean x and covariance P. When using a Kalman filtering method for estimation these parameters are directly provided by the estimation procedure. For other estimation methods, the provided posterior distribution may not be Gaussian (e.g. the Particle Filter) or no distributional information may be provided at all (e.g. Maximum-Likelihood Estimation). In the former case, we can still find a Gaussian distribution that approximates the provided posterior distribution as good as possible. If no distributional information is available at all from the estimation method, the PCRB can be a last resort to find an approximation, assuming that the estimation method's variance is reasonable close to this lower bound.

In chapter 6, the UKF outperforms all other estimation methods, except MLE, for which results are mixed. Since the UKF directly delivers the required Gaussian distribution, we prefer it over MLE and exclusively use it for this chapter's simulation study.

The remainder of this section is organized as follows. First, we present the merging procedure itself, that is how to merge two clusters given the mean and covariance of their parameter estimates. Then, we develop a criterion for determining whether a given merger is beneficial in terms of mean-squared error, by comparing the errors before and after merging. Finally, we describe the hierarchical clustering algorithm based on this criterion.

8.3.1. Merging two Estimation Clusters A and B

Assume we have the demand estimates $(\hat{x}^A, P^A)$ and $(\hat{x}^B, P^B)$ for estimation clusters A and B, respectively. Additionally, both estimates can be decomposed into a volume part $\hat{x}_{vol}$, P_{vol}, a customer choice part $\hat{x}_{choice}$, P_{choice} and the cross-covariance between the two P_{cross} . The simulation uses the Hybrid choice model as in the previous chapters. Here, the price elasticity parameter is the customer choice part $\hat{x}_{choice}$, whereas price-sensitive demand volume and product demand is grouped into the volume part $\hat{x}_{vol}$.

We now consider the joint distribution of both estimates, conditional on $\hat{x}^A_{choice} = \hat{x}^B_{choice}$. That is, we add additional information by assuming that the choice parameters for the two estimation clusters are in fact equal. The resulting conditional distribution is again a multivariate Gaussian distribution, with parameters (see e.g. Kotz et al. 2004):

$$x = \begin{pmatrix} \hat{x}^A_{vol} \\ \hat{x}^B_{vol} \\ \hat{x}^A_{choice} \end{pmatrix} + \begin{pmatrix} -P^A_{cross} \\ P^B_{cross} \\ -P^A_{choice} \end{pmatrix} \cdot \left(P^A_{choice} + P^B_{choice}\right)^{-1} \cdot \left(\hat{x}^A_{choice} - \hat{x}^B_{choice}\right) \tag{8.2}$$

$$\begin{aligned} P = & \begin{pmatrix} P^A_{vol} & 0 & P^A_{cross} \\ 0 & P^B_{vol} & 0 \\ (P^A_{cross})^T & 0 & P^A_{choice} \end{pmatrix} \\ & - \begin{pmatrix} -P^A_{cross} \\ P^B_{cross} \\ -P^A_{choice} \end{pmatrix} \cdot \left(P^A_{choice} + P^B_{choice}\right)^{-1} \cdot \begin{pmatrix} -(P^A_{cross})^T & (P^B_{cross})^T & -P^A_{choice} \end{pmatrix} \end{aligned} \tag{8.3}$$

For subsequent merging steps, only the last block column of P is required, such that only those part needs to be updated in the actual implementation. Although the interpretation is slightly different, equation 8.2 is the straight-forward generalization of the forecast combination procedure of Bates & Granger (1969) to multi-variate forecasts.

Implicitly, the correlation between both estimates is assumed to be zero. This assumption is valid if the observed availabilities and sales are independent for the two estimation clusters. Of course, strictly speaking, network-based availability optimization and customer choice behavior between separate itineraries introduce such dependencies. Especially the feedback loop between optimization and subsequent estimation quality is very complex (see section 4.1.4) and therefore hard to foresee. Thus, our simulation study will have to reveal whether the assumption made here is reasonable.

8.3.2. Comparing the Mean-squared Errors before and after Merging

Equation 8.3 shows that the estimator variance will always be reduced by the merging operation, since all elements on the diagonal of P become smaller. In general however, the true choice parameters x^A_{choice} and x^B_{choice} are not identical. If we produce a single estimate for both, this estimate is necessarily biased when compared to one of the true sets of choice parameters individually. Hence, mean-squared error will reduce during the merging operation if the reduction in variance outweighs the newly introduced bias.

Before any merging operations took place, each estimated parameter vector $\hat{x}^A_{choice}$ corresponds to exactly one true parameter vector x^A_{choice}. This is no longer true, after the first estimation clusters have been merged. Then, the estimated parameter vector

$\hat{x}^A_{choice}$ of a merging cluster corresponds to multiple true parameter vectors, namely to all true parameter vectors of the constituent estimation clusters. We write this set of true parameter vectors as $\{x^i_{choice} : i \in A\}$ with cardinality $|A|$. All of these have to be considered when evaluating the mean-squared error before and after merging..

Formally, the mean-squared error for the choice parameters of estimation cluster A *before* merging is

$$\begin{aligned} MSE^A_{choice,1} &= E\left[\sum_{i\in A}(\hat{x}^A_{choice} - x^i_{choice})^2\right] \\ &= |A| \cdot P^A_{choice} + \sum_{i\in A} E[\hat{x}^A_{choice} - x^i_{choice}]^2. \end{aligned} \tag{8.4}$$

After merging, the mean-squared error for estimation cluster A becomes

$$\begin{aligned} MSE^A_{choice,2} = E&\left[\sum_{i\in A}\left(\hat{x}^A_{choice} - P^A_{choice} \cdot F \cdot (\hat{x}^A_{choice} - \hat{x}^B_{choice}) - x^i_{choice}\right)^2\right] \\ &= MSE^A_{choice,1} + |A| \cdot P^A_{choice} \cdot F \cdot E\left[(\hat{x}^A_{choice} - \hat{x}^B_{choice})^2\right] \cdot F \cdot P^A_{choice} \\ &- 2\sum_{i\in A} E\left[(\hat{x}^A_{choice} - x^i_{choice})(\hat{x}^A_{choice} - \hat{x}^B_{choice})^T\right] \cdot F \cdot P^A_{choice} \end{aligned} \tag{8.5}$$

where $F = \left(P^A_{choice} + P^B_{choice}\right)^{-1}$ and $x^2 = xx^T$. To simplify further, note that

$$E\left[(\hat{x}^A_{choice} - \hat{x}^B_{choice})^2\right] = P^A_{choice} + P^B_{choice} + E\left[\hat{x}^A_{choice} - \hat{x}^B_{choice}\right]^2 \tag{8.6}$$

$$= F^{-1} + E\left[\hat{x}^A_{choice} - \hat{x}^B_{choice}\right]^2 \tag{8.7}$$

and – given $\hat{x}^A_{choice}$ and $\hat{x}^B_{choice}$ are independent – that

$$\begin{aligned} &\sum_{i\in A} E\left[(\hat{x}^A_{choice} - x^i_{choice})(\hat{x}^A_{choice} - \hat{x}^B_{choice})^T\right] \\ &= |A| \cdot P^A_{choice} + |A| \cdot E[\hat{x}^A_{choice}]^2 - |A| \cdot E[(\hat{x}^A_{choice})(\hat{x}^B_{choice})^T] - \sum_{i\in A} x^i_{choice} E[\hat{x}^A_{choice} - \hat{x}^B_{choice}]^T \\ &= |A| \cdot P^A_{choice} + \sum_{i\in A} E[\hat{x}^A_{choice} - x^i_{choice}] E[\hat{x}^A_{choice} - \hat{x}^B_{choice}]^T. \end{aligned} \tag{8.8}$$

Here, we further assume that the estimated parameter vector $\hat{x}^A_{choice}$ is unbiased when compared to the average of all true parameter vectors:

$$E[\hat{x}^A_{choice} - \frac{1}{|A|}\sum_{i\in A} x^i_{choice}] = \frac{1}{|A|}\sum_{i\in A} E[\hat{x}^A_{choice} - x^i_{choice}] = 0 \tag{8.9}$$

In general, this is only approximately true, since the merged estimate is a *weighted* average of the individual estimates and not the simple average required here. Nevertheless, we believe this is a valid approximation here on the basis of our simulation results.

Finally, we have a compact expression for $MSE^A_{choice,2}$:

$$\begin{aligned} MSE^A_{choice,2} = MSE^A_{choice,1} &+ |A| \cdot P^A_{choice} \cdot F \cdot E\left[(\hat{x}^A_{choice} - \hat{x}^B_{choice})^2\right] \cdot F \cdot P^A_{choice} \\ &- |A| \cdot P^A_{choice} \cdot F \cdot P^A_{choice} \end{aligned} \tag{8.10}$$

Now consider the difference $MSE^A_{choice,1} - MSE^A_{choice,2}$ between the two values. If this difference is positive, i.e. a positive definite matrix, then merging the two estimation clusters will reduce the mean-squared error of estimation cluster A. This difference is

$$\begin{aligned} \Delta^A_{MSE} &= MSE^A_{choice,1} - MSE^A_{choice,2} \\ &= |A| \cdot P^A_{choice} \cdot \Big(F - F \cdot E\left[\hat{x}^A_{choice} - \hat{x}^B_{choice}\right]^2 \cdot F\Big) \cdot P^A_{choice} \end{aligned} \tag{8.11}$$

The positive term in equation 8.11 corresponds to variance reduction and the negative term to increased bias. We will use this difference to decide which estimation clusters should be merged.

8.3.3. Initial Clustering

The above discussion suggests the trace of $\Delta^A_{MSE} + \Delta^B_{MSE}$, $tr(\Delta^A_{MSE} + \Delta^B_{MSE})$, as a criterion for clustering, since only a matrix with a positive trace can be positive-definite. The value of this criterion may change after each departure. However, as argued in the beginning of this section, we want to find a stable clustering that can be held fixed over many departures. Therefore we need to know if merging two estimation clusters is beneficial in the long run. To do this, we first replace the covariance matrix P^A_{choice} by the estimated steady-state PCRB (see section 6.4.2) in equation 8.11. Assuming unbiased estimates and an efficient estimation procedure, the steady-state PCRB equals the steady-state covariance matrix.

Additionally, we need a value for $E\left[\hat{x}^A_{choice} - \hat{x}^B_{choice}\right]$. In practice, one would use a secondary source of information to estimate this difference, such as long-term time-series or expert opinions. We observed in section 5.4 that the real-world time series on price sensitivity showed a negative auto-correlation in the long-term. This negative auto-correlation suggests that the time-series is stationary in the long term, which in turn justifies using the long-term average to approximate $E\left[\hat{x}^A_{choice} - \hat{x}^B_{choice}\right]$. This can be combined with expert opinions on the similarity or dis-similarity of markets. Numerous approaches exist in the literature to combine expert forecasts with other, data-driven estimates (for an overview see Clemen 1989). Boyer (2010) explores using various regression techniques to predict willingness-to-pay from overall market properties, such as whether there is low-cost competition, the share of business passengers, etc. However, a detailed treatment of this issue along with a comprehensive empirical study is a research project in itself and outside the scope of this thesis. We do suggest future research on this topic.

In our simulation study, we have no secondary sources of information. Instead we use the true values $x^A_{0,choice}$ and $x^B_{0,choice}$ at the beginning of the simulation and perturb the squared difference by a multiple of the time-evolution covariance matrices Q^A_{choice} and Q^B_{choice}:

$$E\left[\hat{x}^A_{choice} - \hat{x}^B_{choice}\right]^2 = \left(x^A_{0,choice} - x^B_{0,choice}\right)^2 + m \cdot (Q^A_{choice} + Q^B_{choice}). \tag{8.12}$$

The matrices Q^A_{choice} and Q^B_{choice} are those blocks of the time-evolution covariance matrices Q^A and Q^B (see section 4.3), which correspond to the customer choice parameters. Intuitively, equation 8.12 corresponds to a situation, where we knew the exact parameters at $t = 0$ and then these vectors changed by m one-step standard deviations. In the simulation, we use $m = 75$.

With that, the initial clustering algorithm proceeds as follows: first, compute $d^{AB} = tr(\Delta^A_{MSE} + \Delta^B_{MSE})$ for all estimation clusters A and B. Then merge those estimation clusters A and B with the highest d^{AB} to form a merging cluster C, using equations 8.2 and 8.3. With the merged estimates for merging cluster C, compute d^{AC} for all estimation clusters A. Keep merging and updating d until all d's are negative. Then stop and record the resulting clustering hierarchy for use at all future departures.

Our clustering method is an agglomerative hierarchical clustering (see e.g. Hastie et al. 2009, p. 520ff.) with a custom linkage criterion. This linkage criterion is the maximum reduction in mean-squared error. The resulting hierarchy is incomplete, since it stops at some point before creating a single, all-encompassing cluster, namely when no reduction in mean-squared error is possible. It would be feasible to complete the clustering when allowing for increases in mean-squared error; this seems of little practical use though.

8.3.4. Merging Estimates

After each estimation step, the estimates are merged in the order of the clustering hierarchy determined during the initial clustering step. This is done using the parameter values $\hat{x}_{choice}$ and the covariance P_{choice} from the estimation method, again using equations 8.2 and 8.3. The merged estimates are finally used to compute the conditional demand matrices that are passed on to the optimizer. For history building though, the original un-merged estimates are kept, since the merging step might introduce undesirable biases on the estimation cluster level. Feeding those biases back into the estimation algorithm could have undesirable consequences on its performance.

8.4. Simulation Study

Simulation setup is identical to the one in section 7.2.1, however here we additionally implemented our Forecast Merging procedure described above. The results from the previous chapter serve as the benchmark for the results in this chapter. As noted earlier, an initial clustering is performed at the beginning of each simulation using the estimated steady-state PCRB from section 6.4.2 and perturbed demand parameters according to equation 8.12. Then, in each of the 100 runs of a simulation, estimation clusters are merged according to section 8.3.4.

The forecast merging algorithm generated a total of five merging clusters for both scenarios, with 2-12 O&Ds per cluster. Figure 8.3 shows the individual merging steps in the form of dendrograms. Mergers marked by a lower horizontal line happen earlier in the merging process and promise a higher reduction in mean-squared error than mergers with a higher horizontal line. In total, the merging algorithm expected reductions in mean-squared error for the elasticity parameter of 0.101 and 0.142 which is reasonably close to the observed reduction in the simulation of 0.126 and 0.134, respectively. This helps justifying the assumptions made in the derivation of the merging algorithm.

Figure 8.4 shows the percentage change in mean-squared error when forecast merging is applied. The data is split by type of demand parameter and type of trunk leg. The improvement in mean-squared error is most pronounced (more than 95%) for the elasticity parameter. Qualitatively, this is the expected result, since the merging algorithm focuses on that number and because the PCRB suggests that the elasticity parameter

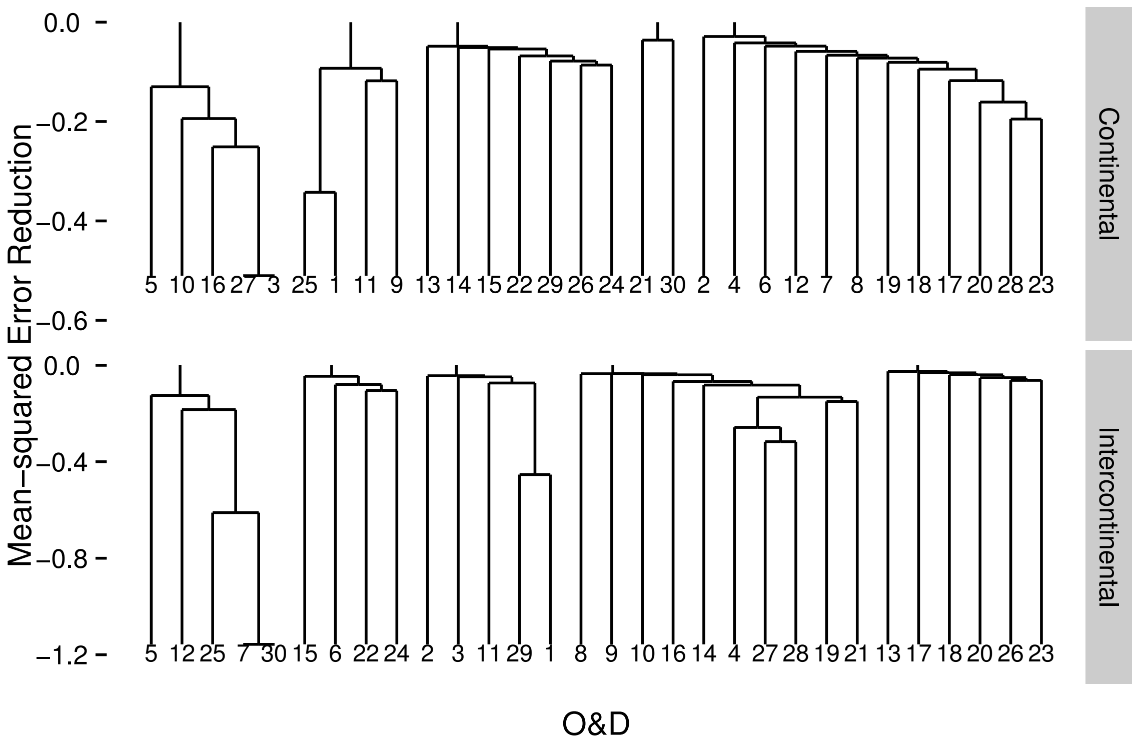

Figure 8.3.: Dendrograms of the forecast merging operation by type of trunk leg; the expected reduction in mean-squared error from the merging operation is shown on the y-axis; O&Ds are numbered by passenger volume.

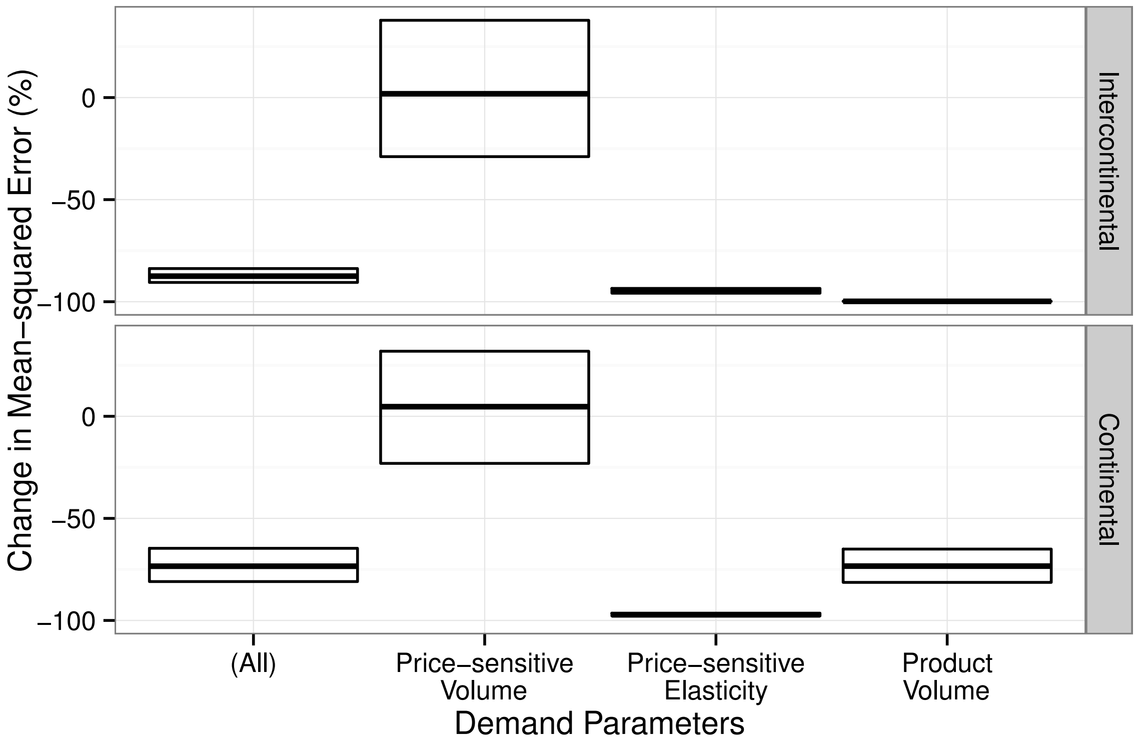

Figure 8.4.: Percentage change in mean-squared error due to forecast merging by type of demand parameter and type of trunk leg; the boxes represent the mean value and its 95% confidence interval.

is the one that can benefit from a larger number of observations.

Interestingly, the product volume parameters also benefit significantly from the merging operation, while price-sensitive volume parameters remain largely unchanged. Correlations in the parameter estimates may lead to adjustments of non-elasticity parameters during forecast merging. If there are strong correlations between two parameters, improving the accuracy of one will also improve the accuracy of the other.

Revenue loss in percent compared to the revenue under a perfect forecast is shown in figure 8.5. Without forecast merging, we observe a revenue loss of about 2%. This number is much higher than the revenue loss from section 6.5.2 where we had only one O&D in the scenario. Without forecast merging, the increased number of O&Ds but same total volume reduces the overall revenue due to the effect of small numbers. With the help of forecast merging, however, revenue loss goes back down to about 0.25%, which is roughly the value observed in section 6.5.2 where data sparsity issues were not present.

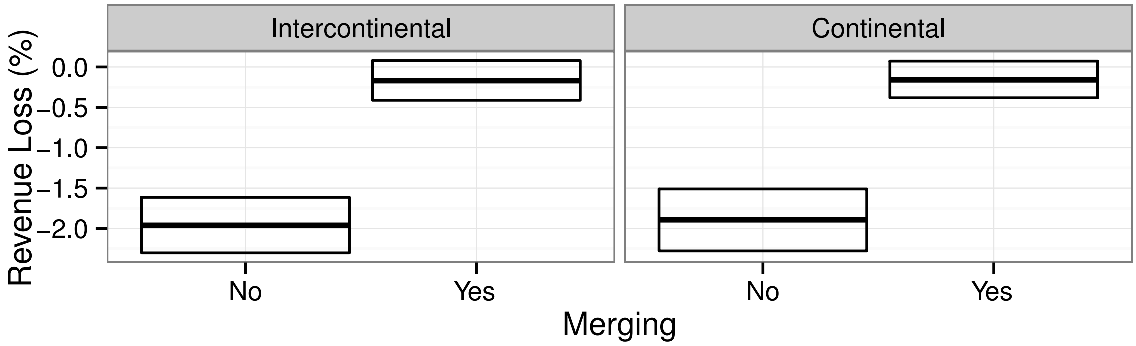

Figure 8.5.: Revenue loss in percent compared to true parameter revenue with and without forecast merging, by type of trunk leg; the boxes represent the mean value and its 95% confidence interval.

Figure 8.6 shows relative estimation errors for the elasticity parameters. The circles and the solid line represent results without forecast merging, while triangles and the dashed line are results with forecast merging. O&Ds are sorted by initial passenger volume. Even though the data is relatively noisy at this level, there is a clear and significant reduction in forecast error over all O&Ds when forecast merging is enabled. This reduction is greatest for the smallest O&Ds on the right, but it also exists for the larger O&Ds to the left. With forecast merging, a small number effect is barely noticeable.

Figure 8.7 plots revenue loss over O&Ds compared to the revenue under a perfect forecast. Again, points and the solid line represent data without forecast merging, while triangles and the dashed line show data with forecast merging. The improved performance due to forecast merging translates from forecast accuracy to revenue. Revenue increases for almost all O&Ds, with the largest relative increases for the smallest O&Ds on the right. The effect of small numbers, so clearly visible without forecast merging, is significantly reduced for the intercontinental scenario and almost eliminated for the continental scenario.

Note however, that care has to be taken interpreting figure 8.7. An improved revenue management system need not necessarily increase revenue on each O&D. In fact, a good O&D revenue management system will deliberately reduce availability and thus revenue for some O&Ds for the benefit of other O&Ds and total revenue. Figure 8.6 is therefore a much better performance indicator on the O&D basis, even if mean-squared forecast error is less tangible than revenue in general.

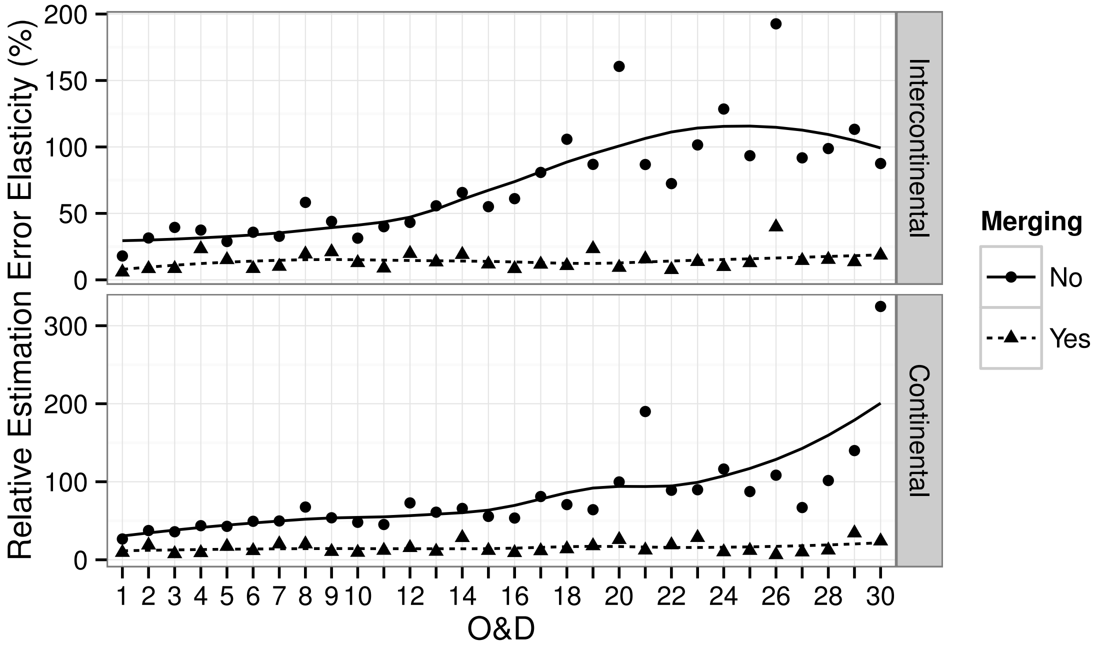

Figure 8.6.: Relative estimation error $\frac{\sqrt{MSE(x)}}{E(x)}$ of elasticity parameters by O&D and with or without forecast merging; O&Ds are sorted by size with the largest O&D, the local O&D, first; points represent the simulation data, averaged over ten independent demand draws, lines are a smoothed estimate of the points and serve as a visual aid.

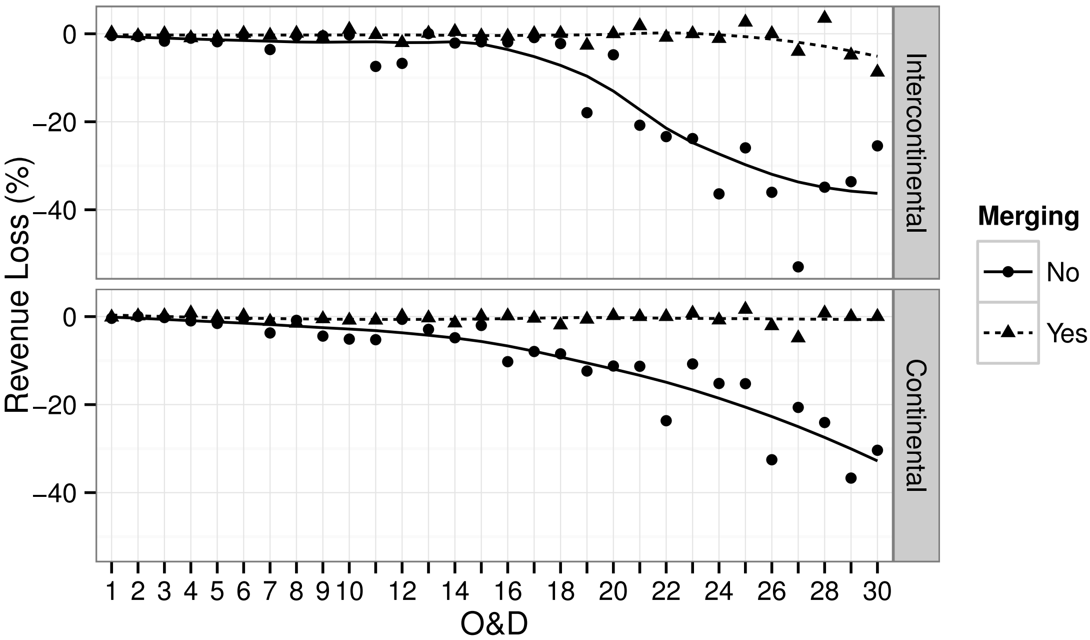

Figure 8.7.: Revenue loss in percent compared to true parameter revenue by O&D, with or without forecast merging; O&Ds are sorted by size with the largest O&D, the local O&D, first; points represent the simulation data, averaged over ten independent demand draws, lines are a smoothed estimate of the points and serve as a visual aid.

8.5. Discussion of Forecast Merging

In both scenarios, forecast merging clearly improved overall revenue with high statistical significance. The average revenue gain is about 1.75%, completely offsetting the negative revenue impact of "small numbers". Reassuringly, revenue increase is highest for small O&Ds and corresponds to similar improvements in forecast accuracy. These results are therefore very promising, demonstrating high potential revenue gains from an easy to implement method that improves the existing forecast system. Boyer (2010) reports revenue gains of about 1%, when moving from separate O&D forecasts to a k-means clustering method. His method does not take estimate uncertainty into account and is not directly minimizing forecast error, which explains why we observe a much larger revenue gain.

Although we do not assume a particular choice model in our derivation, forecast merging requires that the demand parameter vector can be split into a volume and a choice part. This is essential for our method, since we expect only the choice parameters to be similar between different O&Ds and not the volume parameters, according to our findings in chapter 7, in particular figures 7.1 and 7.2. Moreover, motivated by theorem 1 of chapter 7 and the simulation results of section 7.2, we expect forecast merging to only improve the accuracy of choice parameters, since only those are subject to the small number effect. Hence, a demand model where one parameter describes aspects of demand volume and choice simultaneously is not compatible with our forecast merging procedure. The market-sensitive choice model of section 4.1.5 is an example of this.

This problem can be fixed, if a re-parameterization of an incompatible choice model is possible to form an equivalent choice model with distinct parameters for volume and choice. There is another issue however: choice parameters of different O&Ds have to be directly comparable, i.e. similar choice behavior on different O&Ds has to imply similar choice parameters and vice versa. Again, consider the market-sensitive choice model of section 4.1.5 as an example. Here, we can easily imagine a re-parameterization where we extract an overall volume parameter and express the attainable demand and buy-down values as percentages of that overall demand. These modified attainable demand and buy-down values can then legitimately be considered choice parameters. However, these values do not only depend on customer behavior, but also on prices and restrictions. These are rarely the same between separate O&Ds such that the associated choice parameters are not necessarily the same even if customer behavior is identical. In fact, the whole *structure* of the buy-down graph could be different. The problem here is that the market-sensitive choice model does not describe customer choice behavior per se, but the result of customer choice behavior interacting with the airline's offerings. In conclusion, we require a choice model that describes choice behavior itself, such that choice parameters are comparable between O&Ds.

In the simulation, we use the Hybrid choice model of section 4.1.5, which fulfills both criteria: The parameter vector can be split in volume parameters x_{vol}, $x_{product}$ and choice parameters x_{choice}. Moreover, the choice parameter is a price-elasticity parameter, describing the customers reaction to price in relation to a base price. It is therefore not specific to concrete sets of prices, is independent of fare restrictions and even independent of the choice of currency. The choice parameters in the Hybrid choice model are therefore transferable between different O&Ds.

Finally, as noted in the beginning of section 8.3, forecast merging requires the current demand estimate as a multi-variate Gaussian belief distribution. An estimation method that delivers exactly this, is obviously best suited to work in conjunction with forecast merging, and it was such a method, the UKF, that was used in this simulation. The output of other estimation methods can be adjusted accordingly, however, the inaccuracies introduced by this might reduce the benefit of forecast merging.

Part IV.

Conclusion and Appendix

9. Conclusion

9.1. Summary of Findings

The central research objective of this thesis is to investigate the effect of data sparsity on demand estimation in airline revenue management. As a prerequisite, we developed a general model of an airline revenue system in chapter 4 at the beginning of part II of this thesis. It defines the task of each system component and the conceptual data flows between them, yet leaves concrete methods for optimization, demand estimation and availability control, as well as the underlying demand model unspecified. By appending this model with assumptions about demand evolution over time and a probability distribution of bookings, we arrived at a state-space model for the demand estimation problem. This model served as the basis to accurately formulate our ideas and algorithms without the need to assume a specific revenue management methodology from the start. It also let us clearly see and describe the scope of our methods and findings, such as to which general class of choice models they pertain.

In chapter 5, we introduced the simulation setting that we use to evaluate our proposed methods. While complexity is kept as low as possible by restricting the scenario to a single compartment and a single flight, great care has been taken to model the customer choice process as accurately as possible. This has been accomplished by calibrating the parameters of the demand model with real world data and using true fare class prices. This simulation model was used throughout this thesis, however we later extended this model to include a larger flight network with a total of 30 origin-destination pairs. Here, we again used real world data to define the relative market volumes of these 30 connections and the distribution of their price elasticity parameters. Given the strong reliance on actual data, we believe that our simulation setup yields results that can be confidently generalized to real world settings.

Before we were able to zoom in on the issue of data sparsity, we analyzed the problem of demand estimation free from data sparsity issues in chapter 6. These state-space models have been widely studied in the field of control theory. We therefore surveyed available methods from control theory that are applicable to our class of state-space model. Because no exact solution to the general problem with a nonlinear choice function is known, we proposed two heuristics that are applicable in our case and have shown good performance in other applications: the Unscented Kalman Filter and the Particle Filter. With a simulation study we evaluated the performance of these heuristics in a revenue management application, comparing them to benchmark methods and to each other. In a simulation study, we found our proposed methods to clearly outperform two existing methods, Simple Estimation and Forecast Prediction, and to deliver similar results as direct maximization of the likelihood function. Among each other, the UKF lead to higher forecast accuracy and more revenue than the PF. The proposed heuristic for Active Estimation, however, did not show a positive impact on revenues, in fact, the effect was slightly negative in all cases. In conclusion of chapter 6 and part II, we argue for the use of the UKF without Active Estimation, on the basis of these simulation results and auxiliary considerations that make the UKF preferable to MLE.

In part III, we investigated the effect of data sparsity on demand estimation and how it can be avoided. First, chapter 7 answered the question whether a problem of small numbers exists at all and analyzes its structure. Given a demand model that separates between demand volume and customer choice parameters, we showed analytically that the choice parameter forecast error necessarily becomes arbitrarily large when the rate of booking events tends to zero. This result holds regardless of the chosen estimation method. Therefore, this problem of small numbers cannot be alleviated by selecting a different estimation method. Only raising the number of booking events that an estimate is based on can mitigate the negative effect. In that sense, there is a fundamental trade-off between forecast granularity and stability that cannot be broken by finding a better demand estimation method.

While our theoretical result is very general with regard to the set of demand models it applies to and holds for any estimation method, it is only an asymptotic result for the arrival rate tending to zero and furthermore makes no claim on the impact on revenue. We therefore supplemented this result by a simulation study. We observed the expected effect of increasing forecast error for smaller O&Ds, which led to an overall revenue loss of almost 2% compared to 0.3% in chapter 6 where small-number effects were not present.

In conclusion, chapter 7 showed that the problem of small numbers is an inherent property of the demand estimation problem and has the potential for incurring a significant revenue loss in practice. In turn, developing and deploying methods that reduce the effect offers great potential for revenue increases.

Consequently, in chapter 8, we proposed a method to mitigate the impact of data sparsity. Since chapter 7 showed that a different choice of estimation method cannot be the solution and that forecast inaccuracies arise only for customer choice parameters, we proposed a method that merges customer choice information from multiple, similar markets. This increases the effective number of booking events that individual estimates are based on, while keeping separate the estimation of demand volume, which is often vastly different between markets. Our method exploits the fact that our estimation method, the UKF, provides estimation uncertainty information and features a mean-squared error in the same order of magnitude as its theoretical lower bound, the PCRB.

From this, we developed a hierarchical clustering algorithm that automatically stops when no further reduction in forecast error is to be expected. In that way, our Forecast Merging method dynamically adjusts the level of detail for demand estimation, optimizing the trade-off between granularity and stability of the forecast. In a simulation study that uses the same setting as in chapter 7, we showed that Forecast Merging reduces the effect of small numbers to a great extent. This is evident both in forecast error and in revenue with revenue loss back down to the level of chapter 6 where data sparsity issues were not present.

Moreover, it is simple to compute and the clustering hierarchy can be generated offline and inspected manually before implementation. Forecast Merging is therefore a very promising addition to the demand estimation process in practice. However, a simulation study can only ever partially reflect reality. To provide some perspective on the applicability of our results to the real world, we summarized the assumptions and limitations of our simulation study in the following section.

More briefly, the contributions of this thesis can be summarized in the four following items:

1. We formulated the demand estimation and revenue optimization problem as a state-space model. This facilitated the adaptation of state estimation methods, originally developed in control theory, to the airline revenue management problem. (Chapter 4)

2. We showed specifically how two such methods can be adjusted to our problem and compare their performance to existing methods in a simulation study. (Chapter 6)

3. We introduced the "problem of small numbers", which describes the degradation

of forecast accuracy when the number of booking events per estimate becomes smaller and smaller. This is shown to be an inevitable structural property of the estimation problem and observable in a simulation. (Chapter 7)

4. We proposed a forecast merging procedure that mitigates the problem of small numbers by selectively adjusting the level of detail for demand parameter estimation. We evaluated the performance of the proposed procedure in a simulation study. (Chapter 8)

Answering the central research question of this thesis, we showed that revenue management systems that are both network- and choice-based, suffer from a significant negative revenue impact of data sparsity when this issue is not accounted for explicitly. Our proposed Forecast Merging procedure is one such method that addresses data sparsity and can almost completely evade the negative revenue effect.

9.2. Limitations

In this section we review the assumptions and limitations of the simulation studies in this thesis that affect the applicability of our results to real-world revenue management systems. These can be broadly categorized into demand assumptions, supply assumptions, and general limitations of simulation experiments.

Demand Assumptions The simulation employs the Hybrid demand model not only as the customer choice function in forecasting, but also to generate actual demand in the simulation. In reality, demand will not exactly follow the Hybrid demand model, and we chose that model not for its level of realism, but for its simplicity. Using identical demand models in demand generation and forecasting has the advantage of isolating the effect of sub-optimal demand parameter estimation from model misspecification. Furthermore, the “perfect” forecast is clearly defined as the true demand parameters. We relied on this definition throughout this thesis when computing forecast errors and the true parameter revenue.

In a real-world airline revenue management system, the demand model will almost never capture the full complexity of actual demand. In that sense, the demand model will always be mis-specified to some extent. Different estimation methods might be more or less robust towards this. Some methods may find a set of parameters that approximates the true demand as good as possible, but this is certainly not guaranteed.

It may even be true that some estimation methods that perform exceptionally well when true demand follows the model are less robust to model mis-specification. The reason is that these estimation methods heavily exploit the structure of the demand model to achieve their performance under perfect conditions. When this structure is no longer strictly valid, they may suffer more than less sophisticated methods.

This calls for including model mis-specification in a simulation study. Apart from the advantages mentioned above – effect isolation and a defined perfect forecast – there is the issue of not knowing the exact form of model mis-specification. For if the structure of mis-specification was predictable and known, we might just adapt the demand model accordingly. In that sense, we cannot know exactly how true demand differs from our model. Yet, we have to define exactly this to include model mis-specification in the simulation study, and the different choices here may lead to different relative performance of the estimation methods.

Since a demand model can be mis-specified in myriad ways, a comprehensive sensitivity analysis would be extremely time-consuming and is therefore outside the scope of this thesis. Nevertheless, for the reasons mentioned above, it could lead to interesting insights, especially for practitioners. We suggest such a sensitivity analysis for future research.

Apart from defining the demand model itself, its true parameters in the simulation also need to be specified. In a real-world airline network there are thousands of O&D markets with different demand characteristics. Preferably, an estimation method performs well over the complete range. A simulation study should therefore aim to cover this demand variety. It is therefore not a priority to exactly mimic individual real-world markets in the simulation, but instead identify the general range of parameter values and vary the simulation demand in that range. We did just that in sections 5.4 and 7.2.1. As summarized in table 5.1, we varied a number of demand parameters over the three traffic type scenarios "Domestic", "Continental" and "Intercontinental", in addition to overall demand volume. However, from table 5.1 it also becomes evident that many more parameter combinations would be possible, and probably exist in a large airline's network. Again, this is mostly a question of simulation time. Since we did not observe marked differences in estimation performance between our three scenarios, we considered it unlikely that more combinations would create significant additional insight.

Supply Assumptions In contrast to real-world demand, flight schedules and pricing structures are known and publicly posted. Thus, at least in principle, an airline's supply, and that of all its competitors, can be perfectly modeled in a simulation study. Yet, this is not necessarily desirable, as some degree of abstraction and simplification actually increases the utility of a simulation. We discuss the general point in a little more detail below, in the paragraph on general limitations of simulation experiments. In terms of supply, this means that we chose to model the airline's network as small as possible for the research question at hand. In part II, we were mainly interested in demand estimation performance. Since demand estimation is traditionally done separately by O&D, the estimation problems from different O&Ds only interact very loosely through optimization and resulting availabilities. We therefore decided that it suffices to consider only a single O&D and flight in this case. This speeds up the simulation, reduces the amount of produced data and therefore simplifies analysis. Then, in part III, we were explicitly concerned with multiple O&Ds and we therefore had to include them in our simulation. But even here, we limited the number of O&Ds to 30, which is well below the observed number of O&Ds traversing a single flight in the real world. Again, we believe that this number is sufficient to demonstrate the effect of forecast merging, while still keeping simulation time and amount of data to manageable levels.

Pricing structure has also been considerably simplified. While we did use true posted prices, we restricted ourselves to a single compartment, the economy compartment. Furthermore, we modeled the extreme case of no fare restrictions at all. While there is a trend to fewer restriction as argued in section 1, there is still a significant number of fare restrictions present in many airlines' pricing structures.

The strongest assumption on the supply side is the omission of competition. Dennis (2010) analyzes competition in the European airline market, and finds that for non-stop service out of their respective hubs, some network carriers have no competition on up to two thirds of their destinations. Thus, the case of no competition is not completely irrelevant or unrealistic. Yet, on the remaining third of destinations, plus on many connecting markets there is at least one competitor, and other network carriers face much more competition in general. To some degree, a monopolistic revenue management system can handle competition: if competitor behavior is relatively static, then its influence can be included in the customer choice function as a reduced willingness to pay. If the competitor dynamically adjusts its prices however, then the parameters of the customer choice function would have to change accordingly. Unless this happens slowly and gradually, demand estimation from observed data becomes infeasible without explicitly including the competitors behavior in the model. Extending a RM system in such a way is no trivial task though. Apart from the necessity to forecast competitor demand and behavior, game theoretic aspects enter the picture. Competition was therefore outside the scope of this thesis. We refer the reader to Zimmermann (2013) for an in-depth treatment of these issues.

General limitations of simulation experiments Simulations are a widely used and accepted method to evaluate new approaches in revenue management (Frank et al. 2008). However, any simulation is an abstraction and simplification of the real world in many respects. This is generally an advantage, since their greater simplicity facilitates analysis and reveals effects that would otherwise be lost in the noise and complexity of the real world. Therefore, as Frank et al. (2008) argue, a simulation model should only be as accurate as necessary, not as accurate as possible. Finding this level of necessary accuracy, however, is not trivial and requires a combination of good judgement and experience. Even then, there is never a guarantee that some aspect that was judged to be inconsequential and therefore kept out of the simulation, is in fact a significant factor. This is, we believe, an inherent limitation of simulation experiments.

9.3. Directions for Future Research

This thesis focuses on demand estimation when booking data is sparse, abstracting away from other obfuscating factors where possible. This naturally leaves some open questions that stand between our proposed methods and their application in the real world. Furthermore, the simulation setup that we use to obtain or illustrate our results is very specific in the demand model it assumes. Both aspects lead to a number of interesting starting points for future research throughout this thesis which we outline in this section.

Our proposed demand estimation methods in chapter 6 are theoretically applicable to any type of demand function. Our simulation results, however, are based on a particular choice of demand function – the Hybrid Demand function. As such, the conclusions we draw about their relative performance in terms of forecast accuracy and revenue are only valid with respect to the Hybrid Demand function, which we had mostly chosen for its simplicity. A straight-forward extension of our work is therefore to repeat our experiments with other demand functions. Moreover, different estimation methods might suggest themselves for other types of demand functions. Consider e.g. a linear demand model, such as Market-sensitive demand from section 4.1.5, for which the basic Kalman Filter is a natural choice.

Our simulation model is a much simplified image of the real world. While this is desirable when we want to understand the results, and not just observe them, it also means that additional steps have to be taken before our demand estimation methods can be applied to the real world. For example, real world data is often inaccurate, has seasonalities, week-day patterns and does not strictly obey a mathematical demand model. We believe that our state-space formulation of the demand estimation procedure offers a straight-forward path to incorporating these effects in our estimation methods. Instead of using ad-hoc heuristics, data inaccuracies can be treated by adding additional noise terms to the observation equation 4.11; periodic patterns can be explicitly modeled as additional coefficients in the demand function; deviations from the analytical demand model can be partially accounted for as in the Hybrid Demand model, by adding an independent demand function.

At this level of detail, it becomes possible – and probably judicious – to test the demand estimation methods on real world data. Unfortunately, in real world data, the *true* demand parameters are not known. More accurately, they don't even exist, given that real demand does not follow analytical demand models exactly. Yet, comparing actual sales data to expected bookings from the demand estimate does yield a measure of forecast accuracy, the so-called reconstrained forecast error. While this measure can be skewed due to the interdependence between demand estimates and observed availabilities, it still gives an indication of whether a demand estimation method can capture real demand characteristics sufficiently. Additionally, using our knowledge of demand estimate uncertainty and booking variance, the expected value of the reconstrained forecast error can be determined. An observed value that differs significantly from this expectation would be a strong indication for model mismatch.

As mentioned in chapter 2, an additional real-time update procedure could further improve demand estimation, by additionally updating demand parameter estimates during the booking horizon, in the spirit of Lin (2006). This could lead to faster adaptation of the forecast, e.g. in the case of unforeseen events.

In chapter 8, we mentioned that our Forecast Merging procedure requires some initial estimate of customer choice similarities between markets. We suggested to combine offline sources, such as long-term time-series and expert opinions. In our simulation model there are no such outside sources, so we were unable to explore these options in our setting. This is again a case where additional steps are required before our proposed method can be applied in a real world scenario. Moreover, extensions to the demand model, such as seasonalities, should be considered in Forecast Merging as well, since they suffer from small-number effects too, as mentioned in chapter 7. This could be accomplished alongside the choice parameters, or within a separate clustering hierarchy. Again, at this level, it is appropriate to also evaluate the method directly on real world data.

Considering the overall revenue management process, it might be possible to use demand uncertainty information provided by the UKF or PF during optimization. Since revenue optimization is not a linear operation, there is a difference between optimization using a complete belief distribution vs. optimization on the expected values only (Water & Willems 1981). Therefore, an optimization method that takes demand uncertainty explicitly into account might reduce the negative revenue impact of that demand uncertainty.

Including the human revenue management analyst in the picture, an investigation into the value of providing forecast uncertainty information is missing from the literature. We believe that this value could be very high, given that humans are prone to be overconfident in small number settings, and are as such not able to correctly estimate forecast uncertainty intuitively. This has been shown by Tversky & Kahneman (1971) for professional psychologists who should have at least a similar level of statistical training as revenue management analysts.

We developed our methods and ideas in the context of airline revenue management. Concerning the application to other industries, table 1.1 in the introduction shows that our results are most readily transferable to the hotel, rental car, tour operator and passenger railway industries, which fulfill all three of our key assumptions. Still, each of these industries has its own distinct characteristics such that a thorough examination and new simulation studies are needed, before any of our methods can be recommended for implementation in one of these industries.

10. Appendix

10.1. Technical Details for the Unscented Kalman Filter

10.1.1. Upper Triangular Cholesky Decomposition

The standard Cholesky decomposition of a matrix A produces a *lower* triangular matrix G with $GG^T = A$. Define the permutation

$$P = \begin{pmatrix} 0 & 0 & 1 \\ 0 & \iddots & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Since P is a permutation it is ortho-normal such that $P^{-1} = P^T$. Since P is also symmetric, we can further find that $P = P^T = P^{-1}$. Let X be the (regular) Cholesky decomposition of PAP. Then $U = PXP$ is upper triangular and $UU^T = PXPPX^TP = PXX^TP = PPAPP = A$.

10.1.2. Unscented Kalman Filter for Hybrid Choice Models

In this section we derive our formulation of the Unscented Kalman Filter from the equations given by Wan & Merwe (2000).

Let U be the upper triangular Cholesky decomposition of P_t: $P_t = UU^T$. The parameter vector x of length n can be decomposed into a linear part x^L and a non-linear part x^N of length n^L and n^N, respectively, such that $x = (x^L, x^N)^T$. Originally, there are $2n + 1$ sigma points (as opposed to $2n^N + 1$):

$$\sigma_0 = x \tag{10.1}$$

$$\sigma_i = x + \sqrt{n+\kappa}\, U_i \qquad i = 1, \dots, n \tag{10.2}$$

$$\sigma_i = x - \sqrt{n+\kappa}\, U_{i-n} \qquad i = n+1, \dots, 2n \tag{10.3}$$

Each sigma point can be decomposed into its linear the non-linear part:

$$\sigma_i = (\sigma_i^L, \sigma_i^N)^T \tag{10.4}$$

Due to the fact that U is upper triangular, the non-linear part of some of the sigma points remains unchanged: $\sigma_i^N = x^N$ for $i \in I = 0, \ldots, n^L, n+1, \ldots, n+n^L$. Projecting the sigma points through the observation function yields

$$g_i = H_a(\sigma_i) = L_a \sigma_i^L + H_a^N(\sigma_i^N) = g_i^L + g_i^N \qquad i = 0, \ldots, 2n \tag{10.5}$$

The main performance advantage of this formulation comes from a reduction in the number of sigma points that leaves the numerical results unchanged. In this section, we show that it is sufficient to consider σ_0, g_0 and σ_i, g_i for $i \notin I$. The σ_i's and g_i's in the main text are already renumbered, such that only the relevant ones are computed.

Define the weights

$$W_s^0 = \frac{\kappa}{n+\kappa} \tag{10.6}$$

$$W_c^0 = \frac{\kappa}{n+\kappa} + (1 - \alpha^2 + \beta) \tag{10.7}$$

$$W_c^i = W_s^i = \frac{1}{2(n+\kappa)} \qquad i = 1, \ldots, 2n \tag{10.8}$$

Then compute the expected bookings

$$\begin{aligned} z &= \sum_{i=0}^{2n} W_s^i g_i = \sum_{i=0}^{2n} W_s^i (L_a \sigma_i^L + g_i^N) \\ &= L_a \left(\sum_{i=0}^{2n} W_s^i \sigma_i^L \right) + g_0^i \sum_{i \in I} W_s^i + \sum_{i \notin I} W_s^i g_i^N \\ &= L_a x^L + \frac{\kappa + n^L}{\kappa + n} g_0^N + \frac{1}{2(n+\kappa)} \sum_{i \notin I} g_i^N \\ &= z^L + z^N \end{aligned} \tag{10.9}$$

and the booking covariance matrix

$$
\begin{aligned}
P_{zz} &= \sum_{i=0}^{2n} W_c^i (g_i - z)^2 + R = \sum_{i=0}^{2n} W_c^i (g_i^L - z^L + g_i^N - z^N)^2 + R \\
&= \sum_{i=0}^{2n} W_c^i (g_i^L - z^L)^2 + \sum_{i=0}^{2n} W_c^i (g_i^N - z^N)^2 + 2\sum_{i=0}^{2n} W_c^i (g_i^L - z^L)(g_i^N - z^N)^T + R
\end{aligned}
\tag{10.10}
$$

where we use the shorthand $xx^T = x^2$ and where R is the covariance of the error term in the measurement equation, i.e. $R = diag(H_a(x))$. Now consider the individual summands in above equation:

$$
\begin{aligned}
\sum_{i=0}^{2n} W_c^i (g_i^L - z^L)^2 &= L_a \left(\sum_{i=0}^{2n} W_c^i (\sigma_i^L - x^L)^2 \right) L_a^T = L_a \left(\sum_{i=1}^{2n} W_c^i (\sqrt{n+\kappa}\, U_i)^2 \right) L_a^T \\
&= \frac{n+\kappa}{2(n+\kappa)} L_a \left(2 \cdot P^L\right) L_a^T = L_a P^L L_a^T
\end{aligned}
\tag{10.11}
$$

where P^L is the upper left $n^L \times n^L$ block of P_t;

$$
\begin{aligned}
\sum_{i=0}^{2n} W_c^i (g_i^N - z^N)^2 &= \sum_{i \in I} W_c^i (g_i^N - z^N)^2 + \sum_{i \notin I} W_c^i (g_i^N - z^N)^2 \\
&= \left(\frac{n^L + \kappa}{n+\kappa} + (1 - \alpha^2 + \beta) \right) (g_0^N - z^N)^2 + \frac{1}{2(n+\kappa)} \sum_{i \notin I} (g_i^N - z^N)^2
\end{aligned}
\tag{10.12}
$$

$$
\begin{aligned}
2\sum_{i=0}^{2n} W_c^i (g_i^L - z^L)(g_i^N - z^N)^T &= 2L_a \sum_{i \in I} W_c^i (\sigma_i^L - x^L)(g_0^N - z^N)^T \\
&\quad + 2L_a \sum_{i \notin I} W_c^i (\sigma_i^L - x^L)(g_i^N - z^N)^T \\
&= \frac{1}{n+\kappa} L_a \sum_{i \notin I} (\sigma_i^L - x^L)(g_i^N - z^N)^T
\end{aligned}
\tag{10.13}
$$

Finally, consider the demand-booking cross-covariance

$$
\begin{aligned}
P_{xz} &= \sum_{i=0}^{2n} W_c^i(\sigma_i - x)(g_i - z)^T = \sum_{i=0}^{2n} W_c^i(\sigma_i - x)(g_i^L - z^L)^T \\
&+ \sum_{i\in I} W_c^i(\sigma_i - x)(g_0^N - z^N)^T + \sum_{i\notin I} W_c^i(\sigma_i - x)(g_i^N - z^N)^T \\
&= \frac{1}{2(n+\kappa)} \sum_{i=1}^{2n} (\sigma_i - x)(\sigma_i^L - x^L)^T L_a^T + \frac{1}{2(n+\kappa)} \sum_{i\notin I} (\sigma_i - x)(g_i^N - z^N)^T \\
&= P^{NL} \cdot L^T + P_{xz^N}
\end{aligned}
\tag{10.14}
$$

where P^{NL} is the left-most $n \times n^L$ block of P_t. Now all final equations only require σ_0, g_0^N and σ_i, g_i^N for $i \notin I$. Re-indexing i, such that the required indices are in consecutive order, and renaming g^N to g yields the equations from the main text.

10.2. Data Analysis: Additional Charts

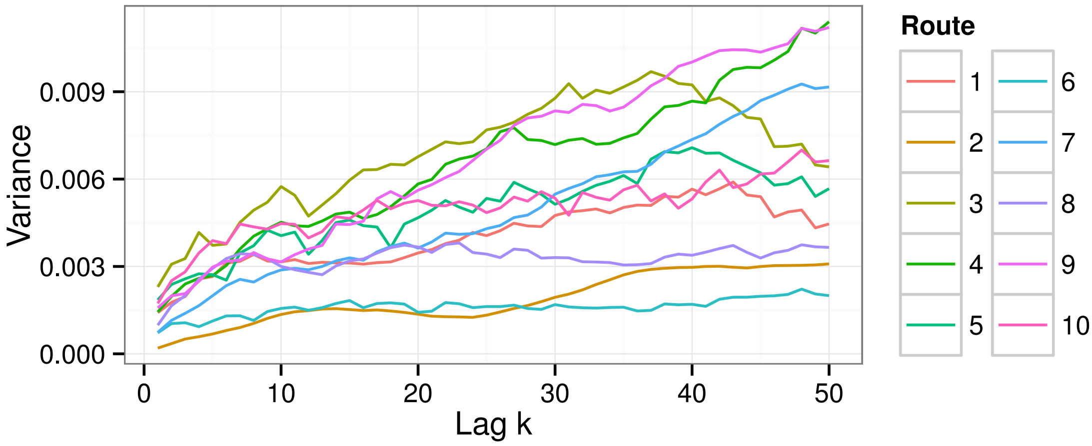

Figure 10.1.: Lag-k variances of elasticity estimate for the ten largest continental routes; $f_{base} = 100$

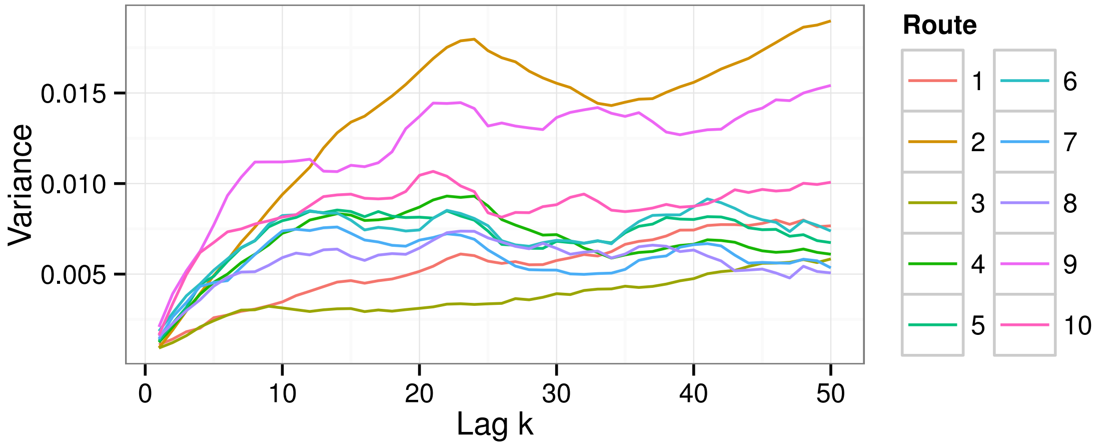

Figure 10.2.: Lag-k variances of elasticity estimate for the ten largest domestic routes; $f_{base} = 100$

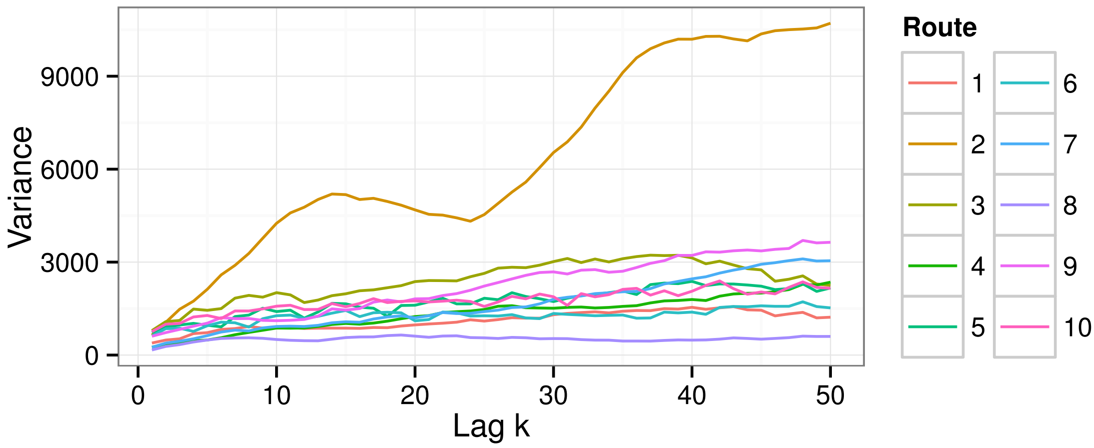

Figure 10.3.: Lag-k variances of volume estimate for the ten largest continental routes; $f_{base} = 100$

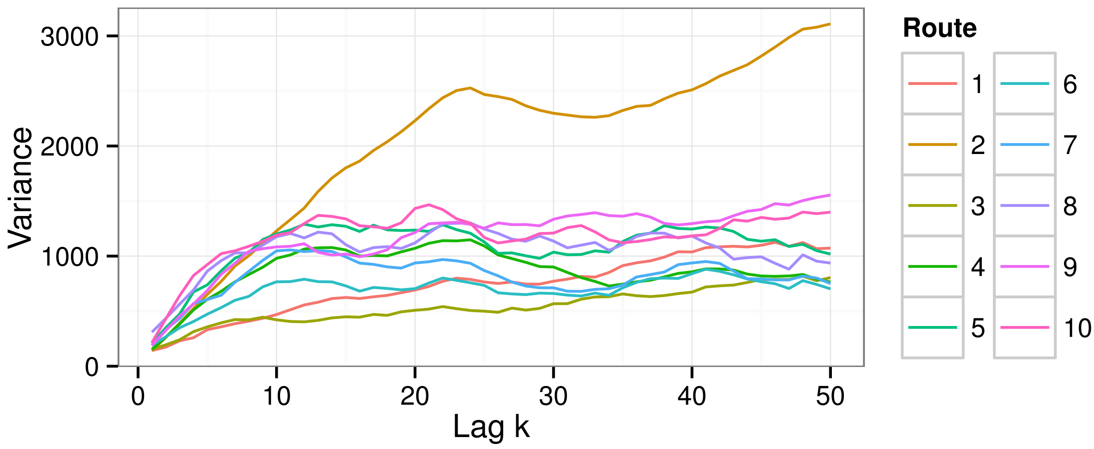

Figure 10.4.: Lag-k variances of volume estimate for then ten largest domestic routes; $f_{base} = 100$

10.3. Fisher Information: Aggregating Observations

In this section we show that the inequality in

$$\begin{aligned} M_{sep} &= E\left[\sum_i \frac{(\nabla_x h_{a_i,i}(x))(\nabla_x h_{a_i,i}(x))^T}{h_{a_i,i}(x)}\right] \\ &\geq E\left[\frac{(\sum_i \nabla_x h_{a_i,i}(x))(\sum_i \nabla_x h_{a_i,i}(x))^T}{\sum_i h_{a_i,i}(x)}\right] = M_{agg} \end{aligned}$$

holds. Let m be the total number of observations and n the number of demand parameters. As usual, only observations with expected bookings $h_{a_i,i}(x) > 0$ are considered. Define the $n \times m$ matrix H as

$$H = \begin{pmatrix} (\nabla_x h_{a_1,1}(x))^T \\ \vdots \\ (\nabla_x h_{a_m,m}(x))^T \end{pmatrix},$$

the diagonal $m \times m$ matrix Δ as

$$\Delta = diag(\frac{1}{h_{a_1,1}(x)}, \ldots, \frac{1}{h_{a_m,m}(x)}),$$

the vector $a = (1, \ldots, 1)^T$ of length m and the sum $\sigma = \sum_{i=1}^m h_{a_i,i}(x)$.

The inequality above holds if the difference between the two Fisher information matrices is positive semi-definite. Using the definitions this difference can be written as

$$\begin{aligned} M_{sep} - M_{agg} &= H^T \Delta H - \frac{1}{\sigma}(a^T H)^T (a^T H) \\ &= H^T (\Delta - \frac{aa^T}{\sigma}) H \end{aligned}$$

It now suffices to show that $(\Delta - \frac{aa^T}{\sigma})$ is positive semi-definite. First, observe that each diagonal element is non-negative, since $\frac{1}{h_{a_i,i}(x)} - \frac{1}{\sigma} \geq 0 \Leftrightarrow h_{a_i,i}(x) \leq \sum_{i=1}^{m} h_{a_i,i}(x)$. Second, each off-diagonal element is smaller in absolute value than the square root of the product of its corresponding diagonal elements. Consider the off-diagonal element at row i and column j:

$$\begin{aligned} & \sqrt{(\frac{1}{h_{a_i,i}(x)} - \frac{1}{\sigma})(\frac{1}{h_{a_j,j}(x)} - \frac{1}{\sigma})} \geq \frac{1}{\sigma} \\ \Leftrightarrow \quad & \sqrt{\frac{1}{h_{a_i,i}(x) \cdot h_{a_j,j}(x)} - \frac{1}{\sigma \cdot h_{a_i,i}(x)} - \frac{1}{\sigma \cdot h_{a_j,j}(x)} + \frac{1}{\sigma^2}} \geq \frac{1}{\sigma} \\ \Leftrightarrow \quad & \frac{1}{h_{a_i,i}(x) \cdot h_{a_j,j}(x)} \geq (\frac{1}{h_{a_i,i}(x)} + \frac{1}{\cdot h_{a_j,j}(x)}) \cdot \frac{1}{\sigma} \\ \Leftrightarrow \quad & h_{a_i,i}(x) + h_{a_j,j}(x) \leq \sum_{i=1}^{m} h_{a_i,i}(x) \end{aligned}$$

Together, these two properties show that $(\Delta - \frac{aa^T}{\sigma})$ has the form of a covariance matrix and is therefore positive semi-definite.

References

Agarwal, D., Broder, A. Z., Chakrabarti, D., Diklic, D., Josifovski, V., & Sayyadian, M. (2007). Estimating Rates of Rare Events at Multiple Resolutions. In *Proceedings of the 13th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 16–25). New York, NY, USA: ACM.

Anandalingam, G., & Chen, L. (1989). Bayesian forecast combination and Kalman filtering. *International Journal of Systems Science*, *20*(8), 1499–1507.

Anderson, E., Bai, Z., Bischof, C., Blackford, S., Demmel, J., Dongarra, J., et al. (1999). *LAPACK Users' Guide* (Third ed.). Philadelphia, PA: Society for Industrial and Applied Mathematics.

Araman, V. F., & Caldentey, R. R. (2009). Dynamic Pricing for Nonperishable Products with Demand Learning. *Operations Research*, *57*(5), 1169–1188.

Åström, K. J., & Murray, R. M. (2008). *Feedback Systems - An Introduction for Scientists and Engineers.* Princeton, New Jersey: Princeton University Press.

Aviv, Y., & Pazgal, A. (2005). *Dynamic Pricing of short life-cycle products through active learning.* Olin School of Business, Washington University, St. Louis, MO. Working Paper. Retrieved from http://apps.olin.wustl.edu/faculty/aviv/papers/Pricing_MS_Dec_2005.pdf

Bartke, P., Cleophas, C., & Zimmermann, B. (2013). Complexity in airline revenue management. *Journal of Revenue and Pricing Management*, *12*(1), 36–45.

Bates, J. M., & Granger, C. W. J. (1969). The Combination of Forecasts. *Operational Research Quarterly*, *20*(4), 451–468.

Belobaba, P. P. (1989). Application of a Probabilistic Decision Model to Airline Seat Inventory Control. *Operations Research*, *37*(2), 183–197.

Besbes, O., & Zeevi, A. (2006). *Blind Nonparametric Revenue Management: Asymptotic Optimality of a Joint Learning and Pricing Method.* Graduate School of Business, Columbia University, NY. Working Paper. Retrieved from http://www.gsb.stanford.edu/sites/default/files/documents/2007_02-14_Besbes.pdf

Bitran, G., & Wadhwa, H. (1996). *A methodology for demand learning with an application to the optimal pricing of seasonal products.* MIT Sloan School of Management, Boston, MA. Working Paper. Retrieved from http://dspace.mit.edu/handle/1721.1/2619

Bobb, Leslie, M., Veral, & Emre. (2008). Open issues and future directions in revenue management. *Journal of Revenue and Pricing Management*, *7*(3), 291–301.

Boyd, E. A., & Kallesen, R. (2004). Practice Papers: The science of revenue management when passengers purchase the lowest available fare. *Journal of Revenue and Pricing Management*, *3*(2), 171–177.

Boyer, C. A. (2010). *Statistical Methods for Forecasting and Estimating Passenger Willingness-to-Pay in Airline Revenue Management.* Master's thesis, Massachusetts Institute of Technology. Retrieved from http://dspace.mit.edu/handle/1721.1/61191

Brankart, J. (2006). *Optimal nonlinear filtering and smoothing assuming truncated gaussian probability distributions in a reduced dimension space of adaptive size* (Tech. Rep.). Grenoble, France: MEOM-LEGI.

Bunn, D., & Topping, I. (1984). Efficiency of the independence assumption in the combination of forecasts. *Operations Research Letters*, *3*(4), 173–178.

Bunn, D. W. (1988). Combining forecasts. *European Journal of Operational Research*, *33*(3), 223–229.

Carvalho, A., & Puterman, M. (2004). *How should a manager set prices when the demand function is unknown.* Statistics Department and Sauder School of Business, University of British Columbia, Vancouver, Canada. Working Paper. Retrieved from http://econpapers.repec.org/paper/ipeipetds/1117.htm

Carvalho, A., & Puterman, M. (2005). Learning and pricing in an internet environment with binomial demands. *Journal of Revenue and Pricing Management*, *3*(4), 320–336.

Chiang, W.-C., Chen, J. C. H., & Xu, X. (2007). An overview of research on revenue management : current issues and future research. *International Journal of Revenue Management*, *1*(1), 97–128.

Chu, L. Y., Shanthikumar, J., & Shen, Z.-J. M. (2008). Solving operational statistics via a Bayesian analysis. *Operations Research Letters*, *36*(1), 110–116.

Chung, B., Li, J., Yao, T., Kwon, C., & Friesz, T. L. (2012). Demand Learning and Dynamic Pricing under Competition in a State-Space Framework. *Engineering Management, IEEE Transactions on*, *59*(2), 240–249.

Cléaz-Savoyen, R. L. (2005). *Airline Revenue Management Methods for Less Restricted Fare Structures.* Master's thesis, Massachusetts Institute of Technology. Retrieved from http://dspace.mit.edu/handle/1721.1/32439

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, *5*(4), 559–583.

Cleophas, C. (2009). *Simulation Based Analysis of Demand Forecast Performance for Airline Revenue Management.* Doctoral dissertation, University of Paderborn. Retrieved from http://nbn-resolving.de/urn:nbn:de:hbz:466-20100303028

Cleophas, C., Frank, M., & Kliewer, N. (2009). Simulation-based key performance indicators for evaluating the quality of airline demand forecasting. *Journal of Revenue and Pricing Management*, *8*(4), 330–342.

Cooper, W. L., Mello, T. Homem-de, & Kleywegt, A. J. A. (2006). Models of the Spiral-Down Effect in Revenue Management. *Operations Research*, *54*(5), 968–987.

Dennis, N. (2010). Old conflicts, new rivals? Airline competition in the European market. In *European transportation conference proceedings.* Henley-in-Arden, UK: Association for European Transport.

Ding, X., Puterman, M. L., & Bisi, A. (2002). The Censored Newsvendor and the Optimal Acquisition of Information. *Operations Research*, *50*(3), 517–527.

Doucet, A., Godsill, S., & Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, *10*(3), 197–208.

Duan, Z., Jilkov, V. P., & Li, X. R. (2008). State estimation with quantized measurements: Approximate MMSE approach. In *11th international conference on information fusion* (pp. 1067–1072). Piscataway, NJ: IEEE.

Duncan, G., Gorr, W., & Szczypula, J. (1993). Bayesian Forecasting for Seemingly Unrelated Time Series: Application to Local Government Revenue Forecasting. *Management Science*, *39*(3), 275–293.

Dunleavy, H., & Phillips, G. (2009). The future of airline revenue management. *Journal of Revenue and Pricing Management*, *8*(4), 388–395.

Easley, D., & Kiefer, N. M. (1988). Controlling a Stochastic Process with Unknown Parameters. *Econometrica*, *56*(5), 1045–1064.

Elliott, G., & Timmermann, A. (2008). Economic Forecasting. *Journal of Economic Literature*, *46*(1), 3–56.

Fiig, T., Isler, K., Hopperstad, C., & Belobaba, P. (2009). Optimization of mixed fare structures: Theory and applications. *Journal of Revenue and Pricing Management*, *9*(1), 152–170.

Fisher, R. A. (1925). Theory of Statistical Estimation. *Mathematical Proceedings of the Cambridge Philosophical Society*, *22*(5), 700–725.

Frank, M., Friedemann, M., & Schroder, A. (2008). Principles for simulations in revenue management. *Journal of Revenue and Pricing Management*, *7*(1), 7–16.

Froemke, C., Hothorn, L., & Schneider, M. (2012). *pairedCI: Confidence intervals for the ratio of locations and for the ratio of scales of two paired samples.*

Gallego, G., & Ryzin, G. van. (1994). Optimal Dynamic Pricing of Inventories with Stochastic Demand over Finite Horizons. *Management Science*, *40*(8), 999–1020.

Gallego, G., & Ryzin, G. van. (1997). A Multiproduct Dynamic Pricing Problem and Its Applications to Network Yield Management. *Operations Research*, *45*(1), 24–41.

Gallego, G., & Talebian, M. (2012). Demand learning and dynamic pricing for multi-version products. *Journal of Pricing and Revenue Management*, *11*(3), 303–318.

Gerlach, M. (2013). *Decentralized Code-Share Revenue Management in Airline Alliances.* Doctoral dissertation, Freie Universität Berlin.

Gordon, N., Salmond, D., & Smith, A. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *Radar and Signal Processing, IEE Proceedings F*, *140*, 107–113.

Guo, J. C. (2008). *Estimation of Sell-up Potential in Airline Revenue Management Systems.* Master's thesis, Massachusetts Institute of Technology. Retrieved from http://dspace.mit.edu/handle/1721.1/45800

Gupta, V., Chung, T. H., Hassibi, B., & Murray, R. M. (2006). On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage. *Automatica*, *42*(2), 251–260.

Gwartney, J. D., Stroup, R. L., Sobel, R. S., & Macpherson, D. A. (2008). *Economics: Private and Public Choice: Private and Public Choice* (12 ed.). Mason, OH: South-Western Cengage Learning.

Handschin, J. E., & Mayne, D. Q. (1969). Monte Carlo techniques to estimate the conditional expectation in multi-stage non-linear filtering. *International Journal of Control*, *9*(5), 547–559.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning* (2nd ed.). New York, NY: Springer.

Julier, S. J. (2002). The scaled unscented transformation. In *Proceedings of the 2002 american control conference* (Vol. 6, pp. 4555–4559). Danvers, MA: American Automatic Control Council.

Julier, S. J., & Uhlmann, J. K. (1997). A New Extension of the Kalman Filter to Nonlinear Systems. In *Proceedings of the conference on signal processing, sensor fusion, and target recognition VI* (pp. 182–193). Bellingham WA: SPIE.

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME-Journal of Basic Engineering*, *82*(Series D), 35–45.

Karlsson, R., & Gustafsson, F. (2005). *Particle Filtering for Quantized Sensor Information* (Tech. Rep.). Linköping, Sweden: Department of Electrical Engineering, Linköping University.

Kiefer, N. M., & Nyarko, Y. (1989). Optimal Control of an Unknown Linear Process with Learning. *International Economic Review*, *30*(3), 571–586.

King, G., & Zeng, L. (2001). Explaining Rare Events in International Relations. *International Organization*, *55*(03), 693–715.

Kitagawa, G. (1996). Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models. *Journal of Computational and Graphical Statistics*, *5*(1), 1–25.

Kleinrock, L. (1975). *Queueing Systems: Theory* (Volume 1 ed.). New York, NY: Wiley-Interscience.

Kong, A., Liu, J. S., & Wong, W. H. (1994). Sequential Imputations and Bayesian Missing Data Problems. *Journal of the American Statistical Association*, *89*(425), 278–288.

Kotz, S., Balakrishnan, N., & Johnson, N. L. (2004). *Continuous Multivariate Distributions, Models and Applications* (2 ed.). New York, NY: John Wiley & Sons Inc.

Kwon, C., Friesz, T. L., Mookherjee, R., Yao, T., & Feng, B. (2009). Non-cooperative competition among revenue maximizing service providers with demand learning. *European Journal of Operational Research*, *197*(3), 981–996.

Lariviere, M. A., & Porteus, E. L. (1999). Stalking Information: Bayesian Inventory Management with Unobserved Lost Sales. *Management Science*, *45*(3), 346–363.

Laub, A. (1979). A Schur method for solving algebraic Riccati equations. *IEEE Transactions on Automatic Control*, *24*(6), 913–921.

Lemke, C., & Gabrys, B. (2008). Forecasting and Forecast Combination in Airline Revenue Management Applications. In *Knowledge processing and reasoning for information society* (pp. 231–247). Warsaw, Poland: EXIT Publishing House.

Lemke, C., Riedel, S., & Gabrys, B. (2013). Evolving forecast combination structures for airline revenue management. *Journal of Revenue and Pricing Management*, *12*(3), 221–234.

Li, J., Yao, T., & Gao, H. (2009). A Revenue Maximizing Strategy Based on Bayesian Analysis of Demand Dynamics. In *SIAM proceedings: Mathematics for industry* (pp. 174–181). San Francisco, CA: Society for Industrial and Applied Mathematics.

Lin, K. Y. (2006). Dynamic pricing with real-time demand learning. *European Journal of Operational Research*, *174*(1), 522–538.

Littlewood, K. (1972). Forecasting and control of passenger bookings. In *Proceedings of the twelfth annual AGIFORS symposium.* New York, NY: American Airlines.

Lobo, M. S., & Boyd, S. (2003). *Pricing and learning with uncertain demand.* Fuqua School of Business, Duke University, Durham, NC. Working Paper. Retrieved from http://www.stanford.edu/~boyd/papers/pric_learn_unc_demand.html

McGill, J. I., & Ryzin, G. J. van. (1999). Revenue Management: Research Overview and Prospects. *Transportation Science*, *33*(2), 233–256.

Ogawa, J. (1983). On the confidence bounds of the ratio of the means of a bivariate normal distribution. *Annals of the Institute of Statistical Mathematics*, *35*(1), 41–48.

Phillips, R. (2005). *Pricing and revenue optimization.* Redwood City, CA: Stanford Business Books.

Price, R. M., & Bonett, D. G. (2000). Estimating the ratio of two Poisson rates. *Computational Statistics & Data Analysis*, *34*(3), 345–356.

R Core Team. (2012). *R: A Language and Environment for Statistical Computing.* Vienna, Austria.

Reyes, M. H. (2006). *Hybrid forecasting for airline revenue management in semi-restricted fare structures.* Master's thesis, MIT. Retrieved from http://dspace.mit.edu/handle/1721.1/34273

Ridout, M., Demétrio, C. G. B., & Hinde, J. (1998). Models for count data with many zeros. In *Proceedings of the XIXth international biometric conference* (pp. 179–192). Cape Town, South Africa: International Biometric Society.

Roff, D. A., & Bentzen, P. (1989). The statistical analysis of mitochondrial DNA polymorphisms: chi 2 and the problem of small samples. *Molecular Biology and Evolution*, *6*(5), 539–545.

Rubin, D. B. (1988). Using the SIR Algorithm to Simulate Posterior Distributions. In J. M. Bernado, D. V. DeGroot, D. V. Lindley, & A. F. M. Smith (Eds.), *Bayesian statistics 3* (Vol. 3, pp. 395–402). Oxford, UK: Oxford University Press.

Schmid, T. (2011). *Spatial Robust Small Area Estimation applied on Business data.* Doctoral dissertation, University of Trier. Retrieved from http://ubt.opus.hbz-nrw.de/volltexte/2012/731/

Şen, A., & Zhang, A. X. (2009). Style goods pricing with demand learning. *European Journal of Operational Research*, *196*(3), 1058–1075.

Shakeri, M., Pattipati, K. R., & Kleinman, D. L. (1995). Optimal measurement scheduling for state estimation. *Aerospace and Electronic Systems, IEEE Transactions on*, *31*(2), 716–729.

Simon, D. (2010). Kalman filtering with state constraints: a survey of linear and nonlinear algorithms. *IET Control Theory & Applications*, *4*(8), 1303–1318.

Smith, B. C., Leimkuhler, J. F., & Darrow, R. M. (1992). Yield Management at American Airlines. *Interfaces*, *22*, 8–31.

Stefanescu, C. (2009). *Multivariate Customer Demand: Modeling and Estimation from Censored Sales.* Management Science and Operations, London Business School, London, United Kingdom. Working Paper. Retrieved from http://papers.ssrn.com/abstract=1334353

Talluri, K. T., & Ryzin, G. van. (2005). *The theory and practice of revenue management.* New York, NY: Springer.

Tichavsky, P., Muravchik, C. H., & Nehorai, A. (1998). Posterior Cramér-Rao Bounds for Discrete-Time Nonlinear Filtering. *IEEE Transactions on Signal Processing*, *46*(5), 1386–1396.

Timmermann, A. (2006). Forecast Combinations. In G. Elliott, C. W. J. Granger, & A. Timmermann (Eds.), *Handbook of economic forecasting* (1st ed., pp. 135–196). Amsterdam: Elsevier.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, *76*(2), 105–110.

Vulcano, G., Ryzin, G. van, & Chaar, W. (2010). OM Practice – choice-based revenue management: An empirical study of estimation and optimization. *Manufacturing & Service Operations Management*, *12*(3), 371–392.

Vulcano, G., Ryzin, G. van, & Ratliff, R. (2012). Estimating Primary Demand for Substitutable Products from Sales Transaction Data. *Operations Research*, *60*(2), 313–334.

Walczak, D. (2006). Modeling high demand variance in dynamic programming. *Journal of Revenue and Pricing Management*, *5*(2), 94–101.

Walczak, D., Mardan, S., & Kallesen, R. (2009). Customer choice, fare adjustments and the marginal expected revenue data transformation: A note on using old yield management techniques in the brave new world of pricing. *Journal of Revenue and Pricing Management*, *9*(1-2), 94–109.

Wan, E. A., & Merwe, R. van der. (2000). The Unscented Kalman filter for nonlinear estimation. In *The IEEE adaptive systems for signal processing, communications, and control symposium 2000* (pp. 153–158). Danvers, MA: IEEE.

Water, H. van de, & Willems, J. C. (1981). The certainty equivalence property in stochastic control theory. *Automatic Control, IEEE Transactions on*, *26*(5), 1080–1087.

Weatherford, L. R. (2009). Survey results from airlines that use revenue management software systems. *Journal of Revenue and Pricing Management*, *8*(4), 323–329.

Weatherford, L. R., Gentry, T. W., & Wilamowski, B. (2003). Neural network forecasting for airlines: A comparative analysis. *Journal of Revenue and Pricing Management*, *1*(4), 319–331.

Weatherford, L. R., & Ratliff, R. M. (2010). Review of revenue management methods with dependent demands. *Journal of Revenue and Pricing Management*, *9*(4), 326–340.

Winkler, R. L., & Makridakis, S. (1983). The Combination of Forecasts. *Journal of the Royal Statistical Society. Series A (General)*, *146*(2), 150–157.

Winter, T. (2010). *Market Sensitive Revenue Optimization: Science and Practice.* Frankfurt am Main, Germany: Presentation at the annual meeting of the GOR Working Group Pricing and Revenue Management.

Winter, T. (2012). *A Combined Dynamic Programming Approach for Bid Price Control of Several Leg Compartments.* Berlin, Germany: Presentation at the INFORMS Revenue Management and Pricing Conference.

Wittenmark, B. (1995). Adaptive dual control methods: An overview. In *5th IFAC symposium on adaptive systems in control and signal processing* (pp. 67–72). New York, NY: Pergamon Press.

Zaritskii, V. S., Svetnik, V. B., & Shimelevich, L. I. (1975). Monte Carlo technique in problems of optimal data processing. *Automation and Remote Control*, *12*, 95–103.

Zimmermann, B. (2013). *Revenue Management with Repeated Competitive Interactions.* Doctoral dissertation, Freie Universität Berlin.

Zimmermann, B., Cleophas, C., & Frank, M. (2011). Universeller Simulator für Revenue Management. *Internationales Verkehrswesen*, *63*(3).